HIERARCHICAL BAYES ANALYSIS FOR CONTINUOUS AND DISCRETE DATA

By

KANNAN NATARAJAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

To my parents and teachers

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Malay Ghosh for being my advisor and for all the attention I received from him for the past four years. Words cannot simply express how grateful I am for his patience, encouragement and invaluable guidance. Without his help it would not have been possible to complete the work. I was extremely fortunate to work under him as his research assistant for most of my years in the graduate program. Also, I would like to thank Professors Alan Agresti, Richard L. Scheaffer, Ramon Littell and Patrick Thompson for their encouragement and advice while serving on my committee.

I am also grateful and highly indebted to my alma mater Indian Statistical Institute, Calcutta, India, for all the support I received. I would like to take this opportunity to thank all the Professors in ISI, to whom I definitely owe a lot for my basic understanding of statistics. I would like to express my appreciation to Professor Bikas K. Sinha for his genuine interest in me. Also, I would like to acknowledge the help and support I received as an undergraduate from the professors in the Department of Statistics, Loyola College, Madras, India.

Much gratitude is owed to my parents, sisters and brothers-in-law, whose support, advice, guidance and prayers throughout the years of my life have made this achievement possible. A Very special thanks are offered to my best friends Atalanta Ghosh and Sofia Paul for their support, friendship and love. There are numerous other friends who made my years in graduate school so memorable and wonderful. Their friendship will never be forgotten.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

HIERARCHICAL BAYES ANALYSIS FOR CONTINUOUS AND DISCRETE DATA

By

Kannan Natarajan

December 1993

Chairman: Malay Ghosh
Major Department: Statistics

This dissertation considers several problems where hierarchical Bayes methodology is used for obtaining estimates and the associated standard errors. Although the methods are presented in the context of some specific problems, they are fairly general in nature and can easily be adapted to other related problems as well.

The first problem is related to the adjustment of census undercount. Adjustment of the decennial census counts in the United States has been a topic of heated debate for more than a decade. Many statisticians, including some within the Bureau of the Census, have recognized the importance of a model based approach for adjustments. In this dissertation, we present a multivariate hierarchical linear model and also relax many of the assumptions which have been the subject of criticism. In particular, we have done a computer-intensive fully Bayesian procedure which uses Monte Carlo numerical integration techniques like the Gibbs sampler. This eliminates the need for

assuming sample variance-covariance matrices of the adjustment factors to be known which was hitherto assumed in any Bayesian or non-Bayesian analysis.

The second specific problem is related to the Quality Measurement Plan (QMP), a plan implemented for reporting the quality assurance audit results to Bell System management. An important function of the Bell Laboratories Quality Assurance Center and the Western Electric Quality Assurance Directorate is to audit the quality of the products manufactured and the services provided by the Western Electric Company to determine if the intended quality standards are met. Starting with the seventh period of 1980, the QMP was implemented. The QMP is based on an empirical Bayes model of the audit-sampling process. It uses the past sample indices but makes an inference about current quality. However, parts of the derivation of QMP are heuristic, including the derivation of the posterior distribution of the current population index, the parameter of interest. Here, we present a hierarchical Bayes model, which avoids the adhoc approximations while deriving the QMP.

The third problem deals with the Bayesian analysis of categorical survey data. Much of the earlier work deals with Bayesian analysis for data in binary fashion, where presence or absence of a specific response is considered. In the case of multi-category data, there will be times, however, when one would like to analyze the responses jointly, arriving at a posterior covariance matrix for the response pattern rather than just a variance for one alternative at a time. Here, a hierarchical Bayesian approach is used to estimate finite population proportions under two-stage sampling within strata based on generalized linear models. In particular, for data on items containing three or more possible responses, a hierarchical Bayesian analysis based on a Poisson model for counts is provided. A Monte Carlo method, the Gibbs sampler, has been used to overcome the computational limitations that have plagued Bayesian analysis for years. The main technique is illustrated using Canada Youth and AIDS Study data.

CHAPTER 1
INTRODUCTION

1.1 Literature Review

Empirical and hierarchical Bayes methods are becoming increasingly popular in statistics, especially in the context of simultaneous estimation of several parameters. For example, agencies of the federal government have been involved in obtaining estimates of per capita income, unemployment rates, crop yields and so forth simultaneously for several state and local government areas. In such situations, quite often estimates of certain area means, or simultaneous estimates of several area means can be improved by incorporating information from similar neighboring areas. Examples of this type are especially suitable for empirical Bayes (EB) analysis. As described in Berger (1985), an EB scenario is one in which known relationships among the coordinates of the parameter vector, say $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)^T$, allow use of the data to estimate some features of the prior distribution. Such problems occur quite frequently in statistics. One such situation is when θ_i arises from some common population; so what we can imagine is creating a probabilistic model for the population and can interpret this model as the prior distribution. For example, one may have reason to believe that the θ_i's are iid from a prior $\pi_0(\lambda)$, where π_0 is structurally known except possibly for some unknown parameter λ. A *parametric empirical Bayes* (EB) procedure is one in which λ is estimated from the marginal distribution of the observations.

Closely related to the EB procedure is the hierarchical Bayes (HB) procedure which models the prior distribution in stages. In the first stage, conditional on $\Lambda = \lambda$, θ_i's are iid with a prior $\pi_0(\lambda)$. In the second stage, a prior distribution (often improper) is assigned to Λ. This is an example of a two stage prior. The idea can be generalized to multistage priors, but will not be pursued in this dissertation.

It is apparent that both the EB and the HB procedures recognize the uncertainty in the prior information. Whereas the HB procedure models the uncertainty in the prior information by assigning a distribution (often *noninformative* or *improper*) to the prior parameters (usually called *hyperparameters*), the EB procedure attempts to estimate the unknown hyperparameters, typically by some classical method such as the method of moments or method of maximum likelihood, etc., and use the resulting estimated priors for inferential purposes. In the context of point estimation, both methods often lead to comparable results. However, when it comes to the question of measuring the standard errors associated with these estimators, the HB method has a clear edge over a naive EB method. Empirical Bayes theory by itself does not indicate how to incorporate the hyperparameter estimation error in the analysis. The HB analysis incorporates such errors automatically and hence is generally more reasonable of the approaches. Also, there are no clear cut measures of standard errors associated with EB point estimators. But the same is not true with HB estimators. To be precise, if one estimates the parameter of interest by its posterior mean, then a very natural estimate of the risk associated with this estimator is its posterior variance. Estimates of the standard errors associated with EB point estimators usually need an *ingenious approximation* (see, e.g., Morris, 1981, 1983), whereas the posterior variances associated with the HB estimators, though often complicated, can be found exactly.

Berger (1985) observes, in addition to the hyperparameter estimation error, two more advantages of using the HB procedure. There are often available both structural

prior information (leading to the first stage prior structure) and subjective prior information about the location of $\boldsymbol{\theta}$. The hierarchical Bayes approach allows the use of both types of information and this can especially be valuable for smaller p. Also, another advantage of the HB approach is that it easily produces more information about the posterior distribution, such as, the posterior covariances, but it would require work to derive in a sophisticated empirical Bayes fashion.

The term *hierarchical Bayes* was first used by Good (1965). Lindley and Smith (1972) called such priors *multistage priors*. The latter used the idea very effectively for estimating the vector of normal means, as well as the vector of regression coefficients. Indeed, Lindley and Smith (1972) reanalyzed the usual linear statistical model using Bayesian methods and the concept of exchangeability. They find estimates in a linear model that substantially improve over the usual estimates derived by the method of least squares, by exploiting the available prior information about the parameters.

There is a huge literature on hierarchical Bayes analysis for a wide range of problems, in the case of continuous data. Much of the literature for continuous data deals with the estimation of parameters of the normal distribution. Ghosh (1992) reviews and unifies the hierarchical and empirical Bayes approach for estimating the multivariate normal mean. To handle the case of heavy tailed priors of the normal distribution, Datta and Lahiri (1992) and Angers and Berger (1991) used t-priors viewing them as scale mixture of normals.

Hierarchical Bayes methodology also have been implemented in improving small area estimators. Empirical Bayes or the variance components approach has been considered for simultaneous estimation of the parameters for several small areas (strata), where each stratum contains a finite number of elements, by Fay and Herriot (1979), Ghosh and Meeden (1986), Ghosh and Lahiri (1987), Battese, Harter and Fuller (1988), and Prasad and Rao (1990). Ghosh and Lahiri (1992) and Datta and Ghosh

(1991) proposed HB procedures as an alternative to the EB procedures for small area estimation problems.

HB procedures also have been used for discrete data in specific contexts. George, Makov and Smith (1992) provide a Bayesian hierarchical analysis of the pump failure data, previously analyzed by Gaver and O'Muircheaetaigh (1987) in an empirical Bayes fashion, by using a Poisson-Gamma hierarchical model. Albert (1988) provides a Bayesian hierarchical generalized linear model (GLM) for the assessment of the goodness of fit of the GLM and the estimation of the mean μ of the random variable from an exponential family. He also discusses tractable accurate approximations for the posterior calculations. The GLM hierarchical model of Albert (1988) is a generalization of the normal hierarchical model of Lindley and Smith (1972). Leonard and Novick (1986) used exchangeable and log-linear hierarchical models for Poisson data while modelling the structure of an r x s contingency table and for drawing marginal inferences about all parameters in the model. Zeger and Karim (1991) cast the generalized linear random effects model in a hierarchical Bayesian framework. The methodology is illustrated through a simulation study and an analysis of infectious disease data by fitting a logistic-normal random effects model.

From a calculational perspective the comparison of the HB approach versus the EB approach previously was something of a toss-up. EB theory requires solving likelihood equations, while the HB approach requires numerical integration, often multi-dimensional. In the past, the use of the HB approach was hampered by the need for multi-dimensional integration. The usual numerical integration tools are not very reliable in high dimensions. Tierney and Kadane (1986), Kass, Tierney and Kadane (1989), Kass and Steffey (1989) have used Laplace's method of approximating marginal posterior densities and moments. The proposed method, like the EB approach, requires solving likelihood equations instead of numerical integrations. In

recent years, with the advent of fast computers, Monte Carlo numerical integration techniques like the Gibbs sampler have become very popular.

By now a large body of literature has evolved dealing with small area estimation problems. One specific problem of small area estimation is related to adjustment of the census undercount. Ericksen and Kadane (1985,1987) proposed a model-based approach toward adjustment of census counts. They advocated shrinking the adjustment factors calculated as the ratio of the 1980 census post enumeration survey (PES) estimates to the census figures toward some suitable regression model similar to the ones considered in Fay and Herriot (1979) and Morris (1983). The model considered by Ericksen and Kadane (1985) is univariate. But Datta et al. (1992), in their article on the 1988 Missouri Dress Rehearsal data discussed a multivariate generalization of the model. These authors developed procedures that were used to model data from the 1990 census and the subsequent PES and smooth survey-based estimates of the adjustment factors.

Hoadley (1981) developed a plan, implemented for reporting the quality assurance audit results to Bell System management, called the Quality Measurement Plan (QMP). The QMP is based on an empirical Bayes model of the audit-sampling process. It represents a considerable improvement in the statistical power for detecting substandard quality as compared with the old rules based on the T-rate system, evolved from the work of Dodge and others. It uses the past indices but makes an inference about current quality.

Another specific problem of small area estimation is related to categorical survey data. Unlike the frequentist approach, the prior structure assumed by the Bayesian approach enables the estimation of the population parameters in cells which contain no data. Stroud (1991) provided a hierarchical-conjugate Bayesian analysis, encompassing simple random, stratified, cluster and two-stage sampling, as well as two-stage

sampling within strata, for data in binary fashion, where presence or absence of a specific response is considered. The main technique was illustrated using a small subset of Canada Youth and AIDS Study data.

1.2 The Subject of this Dissertation

This dissertation considers several problems where hierarchical Bayes methodology is used for obtaining estimates and the associated standard errors. Although the methods are presented in the context of some specific problems, they are fairly general in nature, and can easily be adapted to other related problems as well.

In Chapter 2, we discuss a model-based approach towards adjustment of the 1990 census data. A hierarchical Bayes procedure is proposed, which overcomes many of the criticisms levelled against the Bayesian procedures of earlier authors like Ericksen and Kadane (1985,1987) and Datta et al. (1992). In particular, we have devised a computer-intensive fully Bayesian procedure which uses Monte Carlo numerical integration techniques like the Gibbs sampler. This eliminates the need for assuming sample variance-covariance matrices of the adjustment factors to be known which were hitherto assumed in any Bayesian or non-Bayesian analysis. The findings also indicate that some of the standard errors one obtains by assuming known sample variance-covariance matrices may result in serious underestimation in comparison with what one would have obtained when the uncertainty of such matrices was modelled appropriately.

In Chapter 3, we provide a hierarchical Bayes refinement of Hoadley's Quality Measurement Plan (QMP), which has been severely criticized on several grounds. The HB procedure proposed will avoid the ad hoc approximations needed in Hoadley's original procedure. Also, the method proposed will provide another illustration of the Markov chain Monte Carlo integration technique, Gibbs sampling, which has gained popularity over recent years.

Chapter 4 addresses the Bayesian analysis of categorical survey data, where the data are classified into several (not necessarily two) categories. A hierarchical Bayes procedure is used for the analysis of such data. More generally, a complete HB analysis is given for two-stage sampling within strata based on generalized linear models. The computational limitation of multi-dimensional integration which has plagued Bayesian analysis for years is overcome with the use of the Monte Carlo integration method, the Gibbs sampler. The main technique is illustrated using Canada Youth and AIDS Study data.

CHAPTER 2

ADJUSTMENT OF 1990 CENSUS UNDERCOUNT: A HIERARCHICAL BAYES APPROACH

2.1 Introduction

Adjustment of census counts has been a topic of heated debate for nearly a decade. The 1980 counts were never officially adjusted due to a decision of the then commerce secretary Mr. Robert Moshbacher. However, in several lawsuits brought against the Bureau of the Census by different states and cities who demanded revision of the reported counts, the topic of adjustment came up repeatedly in the courtroom testimony of statisticians appearing as expert witnesses on both sides. The issue was again hotly discussed and debated in subsequent scientific publications (see Ericksen and Kadane, 1985; and Freedman and Navidi, 1986). It is clear from these discussions that the statistics community is sharply divided within itself regarding the desirability of adjusting census counts.

Far from being over, the issue has resurfaced with the appearance of the 1990 census data. Once again secretary Moshbacher announced on July 15, 1991, that the results of the 1990 census would not be adjusted, thus overturning the Census Bureau recommendation to use "adjusted" census data. Almost immediately after this, the city of New York and others brought a lawsuit seeking to overturn the decision of the Commerce Secretary. The case was tried in the courtroom of Federal Judge Joseph M. McLaughlin of Manhattan during May, 1992, and a verdict is yet to come. However,

it is clear that there is yet no concensus even among statisticians on whether or not to adjust the counts.

The objective of this chapter is not to deal with the pros and cons of adjustment but instead to introduce a methodology that can be used for adjustment of the 1990 census if needed. The present method is a refinement and generalizations of the previous work of Datta et al. (1992) where hierarchical and empirical Bayes methods were proposed for adjusting census data. While Datta et al.(1992) analyzed the 1988 Missouri Dress Rehearsal data, the present chapter analyzes the actual 1990 census data.

Like other proponents of adjustment, we agree that the 1990 post enumeration survey (PES) data collected in August 1990 forms the basis of adjustment. The 1990 PES is a sample of 170,000 housing units in 5,400 sample block clusters, each cluster being either one block or a collection of several small blocks. To be useful, the PES results must be generalized to nonsampled blocks. With this end, the population is divided into several groups or poststrata. The census count is known for each such poststratum, while the PES estimates the corresponding true population. The ratio of the PES estimate of the true population to the census count is known as the adjustment factor. The construction of poststrata has undergone several revisions with the original proposal of 1392 postsrata being now replaced by 357 poststrata. The detailed description of the latest poststrata appears in Section 2.3.

We begin at the point where a set of estimated raw adjustment factors and their variances for the different poststrata are available for modelling based on the 1990 census and the subsequent PES. We introduce in Section 2.2 a hierarchical linear model for this purpose, and relax many of the assumptions which have hitherto been the subject of criticism. Ericksen and Kadane (1985) were the first proponents of hierarchical models. Many of the earlier criticisms levelled against their procedure were taken into account in Datta et al. (1992). However, the latter did not model the

sample variance-covariance matrix of the adjustment factors which are estimates and thus bear uncertainty. The present chapter models this uncertainty as well Since the pairwise sample correlation coefficients of the adjustment factors between the different poststrata are much smaller compared to the variances, they are not taken into account in the present analysis.

Section 2.3 contains the actual analysis of the data. We obtain in this section the smoothed adjustment factors and the associated standard errors. We use the Gibbs sampling Monte-Carlo integration technique to carry out the Bayesian analysis. This is in sharp contrast to the previous work of Datta et al. (1992) which used a simple one-dimensional numerical integration subroutine. Due to unknown variances, the present Bayesian analysis involves high-dimensional numerical integration which seems impossible to carry out without resort to some Monte-Carlo integration technique.

There are some important (though not surprising) consequences of the analysis of our data. First, the two sets of point estimates of the true adjustment factors are very close whether we use the present hierarchical Bayes (HB) procedure or the one of Datta et al. (1992). However, for most of the poststrata the standard errors obtained by the present method are 1.5 to 2 times higher than those obtained by the earlier method. From a statistical point of view, this additional variability can be explained very easily from the fact of modelling the sample variance-covariance matrix rather than treating it as fixed. Also, our findings lend some support to Dr. Fay's testimony before Judge McLaughlin (see Fienberg, 1992, p. 35) that the variances of the adjustment factors were understated by a factor of 1.7 to 3.0.

We conclude this section by saying that we believe in the need for adjustment of census data and that a Bayesian analysis is suitable for this purpose. However, to achieve greater robustness, a full HB analysis as done in this chapter is much preferred to a subjective Bayes analysis.

2.2 Hierarchical Bayes Model And Gibbs Sampling

Suppose there are m poststrata. Let Y_i denote the sample adjustment factor for the i th poststratum, and θ_i the corresponding true adjustment factor. Also, let $\hat{V}_i$ denote the sample variance for the i th poststratum ($i = 1, \cdots, m$).

The following HB model is considered.

I. Conditional on $\theta_1, \cdots, \theta_m, V_1, \cdots, V_m, \boldsymbol{\beta}$ and σ^2, Y_i and $\hat{V}_i$'s are mutually independent with $Y_i \overset{\text{ind}}{\sim} N(\theta_i\ ,\ V_i)$ and $\hat{V}_i \overset{\text{ind}}{\sim} V_i \frac{\chi^2_{n_i}}{n_i}$;

II. $\theta_1, \cdots, \theta_m \mid V_1, \cdots, V_m, \boldsymbol{\beta}, \sigma^2 \overset{\text{ind}}{\sim} N(\boldsymbol{x}_i^T \boldsymbol{\beta}\ ,\ \sigma^2)$;

III. Marginally, $\boldsymbol{\beta}, \sigma^2, V_1, \cdots, V_m$ are mutually independent with $\boldsymbol{\beta} \sim$Uniform($\boldsymbol{R}^p$). Z=$(\sigma^2)^{-1}$ $\sim$ Gamma($\frac{1}{2}$c , $\frac{1}{2}$d) and $\xi_i = V_i^{-1}$'s are Gamma($\frac{1}{2}$a , $\frac{1}{2}$b). [A random variable W is said to have a Gamma(α, β) distribution if it has a pdf of the form f(w) $\propto exp(-\alpha w) w^{\beta-1}\ I_{(0,\infty)}(w)$, where I denotes the usual indicator function]. We allow the possibility of diffuse priors for Z or ξ_i's, for example a=c=0, etc. Note that the above hierarchical model is suitable for other contexts as well, for example in the estimation of income of small places as considered by Fay and Herriot(1979). These authors considered an alternative empirical Bayes approach for this problem.

We shall use the notations $\boldsymbol{Y} = (Y_1, \cdots, Y_m)^T$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$, $\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)$. Then the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$=$\boldsymbol{y}$ and $\hat{V}_i = \hat{v}_i$ ($i = 1, \cdots, m$) is obtained as follows:

(i) conditional on $Y_i = y_i, \hat{V}_i = \hat{v}_i$, $\xi_i (i = 1, \cdots, m)$ and Z = z, $\boldsymbol{\theta} \sim \boldsymbol{N}_m(\boldsymbol{E}^{-1}\boldsymbol{\Lambda}\boldsymbol{y}, \boldsymbol{E}^{-1})$, where $\boldsymbol{\Lambda}$ = diag$(\xi_1, \cdots, \xi_m)$ and $\boldsymbol{E} = \boldsymbol{\Lambda} + z(\boldsymbol{I}_m - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)$;

(ii) conditional on $Y_i = y_i$ and $\hat{V}_i = \hat{v}_i (i = 1, \cdots, m)$, Z, $\xi_1, \cdots, \xi_m$ have joint pdf

$$f(z, \xi_1, \cdots, \xi_m \mid y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m) \propto |\ \boldsymbol{E}\ |^{-1/2}\ exp\left[-\frac{1}{2}\boldsymbol{y}^T\left(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{E}^{-1}\boldsymbol{\Lambda}\right)\boldsymbol{y}\right]$$

$$exp\left[-\frac{1}{2}cz\right]\ z^{\frac{1}{2}(m-p+d)-1}$$

$$\prod_{i=1}^{m}\left\{\xi_i^{\frac{1}{2}(n_i+b-1)}exp\left(-\frac{1}{2}\xi_i(n_i\hat{v}_i+a)\right)\right\}$$

.

Finding the posterior distribution of $\boldsymbol{\theta}$ through (i) and (ii) requires evaluation of (m+1)-dimensional integrals. The task becomes quite formidable even for moderate m. Rather than using multidimensional numerical integral, we use Monte Carlo numerical integration to generate the posterior distributions and associated means and variances. More specifically, we use Gibbs sampling originally introduced in Geman and Geman(1984), and more recently popularized by Gelfand and Smith(1990) and Gelfand et al. (1990). Gibbs sampling is described below.

Gibbs sampling is a Markovian updating scheme. Given an arbitrary starting set of values $U_1^{(0)},\cdots,U_k^{(0)}$, we draw $U_1^{(1)} \sim [U_1 \mid U_2^{(0)},\cdots,U_k^{(0)}]$, $U_2^{(1)} \sim [U_2 \mid U_1^{(1)},U_3^{(0)},\cdots,U_k^{(0)}],\cdots,U_k^{(1)} \sim [U_k \mid U_1^{(1)},\cdots,U_{k-1}^{(1)}]$, where $[\cdot \mid \cdot]$ denotes the relevant conditional distributions. Thus, each variable is visited in the natural order and a cycle in this scheme requires k random variate generations. After t such iterations, one arrives at $(U_1^{(t)},\cdots,U_k^{(t)})$. As t$\rightarrow\infty$, $(U_1^{(t)},\cdots,U_k^{(t)}) \xrightarrow{d} (U_1,\cdots,U_k)$. Gibbs sampling through q replications of the aforementioned t-iterations generates q iid k-tuples $(U_{1j}^{(t)},\cdots,U_{kj}^{(t)})$ (j =1,$\cdots$,q). $U_1,\cdots,U_k$ could possibly be vectors in the above scheme.

Using Gibbs sampling, the joint posterior pdf of $\theta_1,\cdots,\theta_m$ is approximated by

$$q^{-1}\sum_{j=1}^{q}\left[(\theta_1,\cdots,\theta_m)\mid \boldsymbol{y},\hat{v}_1,\cdots,\hat{v}_m,\boldsymbol{\beta}=\boldsymbol{\beta}_j^{(t)},z=z_j^{(t)},\boldsymbol{v}_i=\boldsymbol{v}_{ij}^{(t)},i=1,\cdots,m\right] \quad (2.2.1)$$

To estimate the posterior moments, we use Rao-Blackwellized estimates as in Gelfand and Smith (1991). Notice that

$$\boldsymbol{E}(\theta_i \mid \boldsymbol{y},\hat{v}_i,\cdots,\hat{v}_m,\xi_1,\cdots,\xi_m,\boldsymbol{\beta},z)=(1-B_i)y_i + B_i\boldsymbol{x}_i^T\boldsymbol{\beta},$$

where $B_i = z/(z+\xi_i)$, $i = 1, \cdots, m$. This is approximated by

$$q^{-1} \sum_{j=1}^{q} \left((1 - B_{ij}^{(t)}) y_i \; + \; B_{ij}^{(t)} \boldsymbol{x}_i^T \boldsymbol{\beta}_j^{(t)} \right), \tag{2.2.2}$$

where as before, t denotes the number of iterations needed to generate a sample. Next noting that

$$\begin{aligned}
&V(\theta_i \mid y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m) \\
&= E[\, V(\theta_i \;\mid\; y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m, \xi_1, \cdots, \xi_m, \boldsymbol{\beta}, z) \mid y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m] \\
&+ V[\, E(\theta_i \;\mid\; y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m, \xi_1, \cdots, \xi_m, \boldsymbol{\beta}, z) \mid y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m] \\
&= E[\, (z + \xi_i)^{-1} \;\mid\; y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m] \\
&+ V[\, (1 - B_i) y_i + B_i \boldsymbol{x}_i^T \boldsymbol{\beta} \;\mid\; y_1, \cdots, y_m, \hat{v}_1, \cdots, \hat{v}_m] \,,
\end{aligned} \tag{2.2.3}$$

one approximates the same by

$$q^{-1} \sum_{j=1}^{q} \left(z_j^{(t)} + \xi_{ij}^{(t)} \right)^{-1} + q^{-1} \sum_{j=1}^{q} B_{ij}^{(t)^2} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_j^{(t)} \right)^2 - \left[q^{-1} \sum_{j=1}^{q} B_{ij}^{(t)} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_j^{(t)} \right) \right]^2 \tag{2.2.4}$$

The Gibbs sampling analysis is based on the following posterior distributions :

(i) $\boldsymbol{\beta} \mid \boldsymbol{y}, \hat{v}_1, \cdots, \hat{v}_m, \boldsymbol{\theta}, \xi_1, \cdots, \xi_m, z \sim \boldsymbol{N}_p \left((\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\theta} \,,\, z^{-1} (\boldsymbol{X}^T \boldsymbol{X})^{-1} \right)$;

(ii) $z \mid \boldsymbol{y}, \hat{v}_1, \cdots, \hat{v}_m, \boldsymbol{\beta}, \boldsymbol{\theta}, \xi_1, \cdots, \xi_m \sim \text{Gamma}\left(\frac{1}{2}(c + \sum_{i=1}^{m} (\theta_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2) \,,\, \frac{1}{2}(d + m) \right)$;

(iii) $\xi_1, \cdots, \xi_m \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\beta}, z, \hat{v}_1, \cdots, \hat{v}_m \overset{ind}{\sim} \text{Gamma}\left(\frac{1}{2}(a + (y_i - \theta_i)^2 + n_i \hat{v}_i) \,,\, \frac{1}{2}(n_i + b + 1) \right)$;

(iv) $\theta_1, \cdots, \theta_m \mid \boldsymbol{y}, \hat{v}_1, \cdots, \hat{v}_m, \boldsymbol{\beta}, z \overset{ind}{\sim} N \left((1 - B_i) y_i + B_i \boldsymbol{x}_i^T \boldsymbol{\beta} \,,\, z^{-1} B_i \right)$.

We investigate in the next section how this approach leads to smoothed adjustment factors for the 1990 census.

2.3 Adjustment of 1990 Census Data

As mentioned in the introduction, the latest adjustment factors are available for 357 poststrata. We now give a brief description of what these poststrata are. First, for non-Hispanic white and other owners, there are four geographic areas under consideration: (i) northeast, (ii) south, (iii) midwest and (iv) west. Each geographic area is then divided into (a) urbanized areas with population 250,000+, (b) other urban areas, and (c) nonurban area. This leads to 12 strata. Similarly for non-Hispanic white and other non-owners (renters), there are 12 such strata. Next black owners in urbanized areas with population 250,000+ are classified into four strata according to four geographic areas. However, black owners in other urban areas are collapsed into one stratum as are black owners in nonurban areas. This leads to 6 strata for black owners. Similarly each category of black nonowners, nonblack Hispanic owners, and nonblack Hispanic nonowners is divided into 6 strata following the same pattern used in the construction of strata for the black owners.

So far we have reached a total of 12+12+6+6+6+6 = 48 strata. Added to these are 3 strata containing (i) Asian and Pacific-Islander owners, (ii) Asian and Pacific-Islander nonowners, and (iii) American Indians on reservations. This leads to a total of 51 strata. Each such stratum is now cross-classified with 7 age-sex categories: (a) 0-17 (males and females), (b) 18-29 (males), (c) 18-29 (females), (d) 30-49 (males), (e) 30-49 (females), (f) 50+ (males), and (g) 50+ (females). This leads to a total of 51 x 7 = 357 poststrata.

The set of adjustment factors and the sample variances are available for all the 357 poststrata. However, for performing the HB analysis, we have not taken into account the last three categories of (i) Asian and Pacific-Islander owners, (ii) Asian and Pacific-Islander nonowners, and (iii) American Indians on reservations, as it is generally felt that these categories should not be merged with the rest, and an

HB analysis combines information from all the sources in calculating the smoothed adjustment factors. This leads to a HB analysis based on 336 poststrata. We do not report that analysis here but discuss instead the results of a simpler analysis based on 48 poststrata where all seven age-sex categories are pooled into one. Even with this simplification, the main messages of this chapter, namely the need for (i) smoothing the adjustment factors and (ii) providing more reliable estimates of the associated standard errors, are clearly conveyed in our analysis.

We consider the hierarchical model as given in Section 2.2 with a=b=c=d=0 to ensure some form of diffuse gamma priors for the inverse of the variance components in our model. The results, however, are not very sensitive to the choice of a, b, c, d as long as some version of diffuse prior is used. Next, the n_i's, the degrees of freedom for the χ^2 distribution associated with $\hat{V}_i$ in the i th poststratum, represent the P - sample (the number of persons counted in the PES) in the i th poststratum divided by some factor, here 300 We admit the adhockery of the number 300, but feel that division by some such factor is essential to perform some meaningful analysis. The design matrix $\boldsymbol{X}$ provided to us from the Bureau of the Census was obtained via best subsets regression and is of the form

$$\boldsymbol{X}^T = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{48}) ,$$

where each $\boldsymbol{x}_i$ is a nine component column-vector with the first element equal to 1, the second element equal to the indicator for nonowner, the third and the fourth elements equal to the indicators for black and Hispanic respectively, the fifth and the sixth elements denoting, respectively, the indicators for an urbanized area with a population of 250,000+ and a nonurbanized area, and finally the seventh, eighth and ninth elements denoting, respectively, the indicator or proportion in northeast, south and west.

The HB analysis was performed by using the Gibbs sampler. In performing the analysis, we have taken t (the number of iterations needed to generate a sample)

equal to 50, while the number of samples is taken as 2500. The stability in the point estimates of the adjustment factors is achieved once a sample of 1500 is generated while stability in the associated standard errors is achieved once a sample of 2500 is generated.

The results of the HB analysis are reported in Table 2.1 which provides the adjustment factors (Y), the corresponding standard errors (SD.Y), the smoothed adjustment factors using the hierarchical model of Section 2.2 (HB1), the associated standard errors (SD.HB1), the smoothed adjustment factors using the model of Datta et al. (HB2) and the associated standard errors (SD.HB2) for all the 48 poststrata

It is clear from Table 2.1 that both the present method and the one of Datta et al. essentially lead to the same point estimates of the adjustment factors and both the methods lead to substantial reduction in the standard errors. However, in most of the 48 poststrata, the estimated standard errors obtained by the present method (SD.HB1) are 1.5 to 2 times (sometimes even more) bigger than the ones of Datta et al. (SD.HB2). A few exceptions are poststrata 12, 25, 27, 28, 31-34 where the estimated standard errors using the present method are lower than the ones using the model of Datta et al. This is somewhat surprising, and we do not have an intuitive explanation of this phenomenon as yet.

We conclude with the assertion that a model-based approach for smoothing the adjustment factors is strongly recommended. Also, hierarchical modelling is particularly well-suited to meet this need.

TABLE 2.1. RAW ADJUSTMENT FACTORS, HB ESTIMATORS AND STANDARD ERRORS

I	Y	SD.Y	HB1	SD.HB1	HB2	SD.HB2
1	0.9792	0.0104	0.9902	0.0038	0.9897	.0027
2	1.0069	0.0072	1.0038	0.0042	1.0030	.0020
3	0.9974	0.0039	0.9948	0.0023	0.9949	.0015
4	0.9966	0.0064	1.0027	0.0034	1.0054	.0024
5	0.9893	0.0048	0.9908	0.0030	0.9909	.0021
6	1.0052	0.0043	1.0044	0.0034	1.0041	.0023
7	0.9990	0.0040	0.9954	0.0024	0.9961	.0020
8	1.0063	0.0058	1.0035	0.0026	1.0067	.0029
9	0.9947	0.0069	0.9937	0.0046	0.9926	.0025
10	1.0018	0.0069	1.0072	0.0043	1.0057	.0032
11	0.9930	0.0116	0.9981	0.0048	0.9975	.0032
12	1.0029	0.0069	1.0063	0.0034	1.0083	.0037
13	1.0117	0.0143	1.0238	0.0062	1.0272	.0036
14	1.0262	0.0156	1.0374	0.0045	1.0404	.0025
15	1.0239	0.0170	1.0283	0.0046	1.0322	.0027
16	1.0328	0.0172	1.0365	0.0055	1.0430	.0025
17	1.0353	0.0162	1.0245	0.0060	1.0284	.0029
18	1.0330	0.0186	1.0380	0.0040	1.0415	.0025
19	1.0124	0.0113	1.0288	0.0048	1.0332	.0028
20	1.0470	0.0147	1.0371	0.0052	1.0442	.0026
21	1.0697	0.0467	1.0274	0.0079	1.0301	.0029
22	1.0665	0.0193	1.0409	0.0060	1.0433	.0029
23	1.0293	0.0160	1.0318	0.0073	1.0350	.0030
24	1.0648	0.0206	1.0400	0.0068	1.0459	.0033

Table 2.1 (continued)

I	Y	SD.Y	HB1	SD.HB1	HB2	SD.HB2
25	1.0165	0.0196	1.0108	0.0057	1.0071	.0084
26	1.0221	0.0094	1.0242	0.0065	1.0202	.0065
27	1.0082	0.0088	1.0151	0.0058	1.0120	.0066
28	1.0649	0.0216	1.0234	0.0058	1.0230	.0063
29	1.0136	0.0101	1.0230	0.0067	1.0198	.0055
30	1.0364	0.0203	1.0271	0.0084	1.0226	.0051
31	1.0913	0.0193	1.0445	0.0075	1.0448	.0095
32	1.0669	0.0217	1.0579	0.0067	1.0578	.0075
33	1.0638	0.0191	1.0489	0.0070	1.0496	.0078
34	1.1106	0.0335	1.0570	0.0071	1.0604	.0072
35	1.0433	0.0128	1.0561	0.0071	1.0568	.0065
36	1.0484	0.0595	1.0605	0.0094	1.0598	.0059
37	1.0068	0.0444	1.0132	0.0052	1.0107	.0047
38	1.0259	0.0095	1.0267	0.0046	1.0239	.0030
39	0.9585	0.0238	1.0176	0.0049	1.0156	.0036
40	1.0298	0.0092	1.0258	0.0046	1.0265	.0030
41	1.0095	0.0170	1.0255	0.0044	1.0249	.0029
42	1.0280	0.0283	1.0282	0.0053	1.0262	.0030
43	1.0721	0.0404	1.0469	0.0075	1.0483	.0055
44	1.1030	0.0311	1.0604	0.0054	1.0614	.0039
45	1.0711	0.0374	1.0513	0.0066	1.0532	.0043
46	1.0629	0.0209	1.0595	0.0067	1.0640	.0036
47	1.0707	0.0310	1.0584	0.0062	1.0618	.0032
48	1.1876	0.0724	1.0621	0.0077	1.0644	.0032

Table 2.2 HB1 and HB2

I	HB1	SD.HB1	HB2	SD.HB2
1	0.9902	.00384	0.9897	.00274
2	1.0038	.00421	1.0030	.00195
3	0.9948	.00232	0.9949	.00145
4	1.0027	.00342	1.0054	.00238
5	0.9908	.00297	0.9909	.00214
6	1.0044	.00344	1.0041	.00231
7	0.9954	.00244	0.9961	.00195
8	1.0035	.00263	1.0067	.00288
9	0.9937	.00464	0.9926	.00245
10	1.0072	.00426	1.0057	.00323
11	0.9981	.00481	0.9975	.00315
12	1.0063	.00344	1.0083	.00373
13	1.0238	.00624	1.0272	.00357
14	1.0374	.00449	1.0404	.00251
15	1.0283	.00457	1.0322	.00266
16	1.0365	.00550	1.0430	.00245
17	1.0245	.00598	1.0284	.00288
18	1.0380	.00403	1.0415	.00249
19	1.0288	.00476	1.0332	.00276
20	1.0371	.00520	1.0442	.00262
21	1.0274	.00786	1.0301	.00291
22	1.0409	.00599	1.0433	.00292
23	1.0318	.00726	1.0350	.00297
24	1.0400	.00677	1.0459	.00327

Table 2.2 (continued)

25	1.0108	.00565	1.0071	.00839
26	1.0242	.00649	1.0202	.00649
27	1.0151	.00580	1.0120	.00660
28	1.0234	.00578	1.0230	.00634
29	1.0230	.00671	1.0198	.00551
30	1.0271	.00836	1.0226	.00505
31	1.0445	.00754	1.0448	.00948
32	1.0579	.00665	1.0578	.00754
33	1.0489	.00697	1.0496	.00777
34	1.0570	.00714	1.0604	.00722
35	1.0561	.00712	1.0568	.00647
36	1.0605	.00935	1.0598	.00592
37	1.0132	.00524	1.0107	.00465
38	1.0267	.00463	1.0239	.00298
39	1.0176	.00492	1.0156	.00357
40	1.0258	.00461	1.0265	.00297
41	1.0255	.00439	1.0249	.00285
42	1.0282	.00526	1.0262	.00304
43	1.0469	.00752	1.0483	.00550
44	1.0604	.00541	1.0614	.00385
45	1.0513	.00660	1.0532	.00430
46	1.0595	.00667	1.0640	.00358
47	1.0584	.00619	1.0618	.00324
48	1.0621	.00770	1.0644	.00316

CHAPTER 3

REFINEMENT OF QUALITY MEASUREMENT PLAN

3.1 Introduction

The primary responsibility of Bell Laboratories Quality Assurance Center (QAC) is to maintain quality requirements in the communication products designed by Bell Laboratories, manufactured by Western Electric Company, Incorporated, and then marketed to Bell System operating companies. In order to meet this responsibility, the QAC conducts quality assurance audits on the products along with its Western Electric agents, the Quality Assurance Directorate (QAD) and Purchased Products Inspection (PPI) organizations.

Quality assurance audits are a structured system of inspections done on a sampling basis by inspectors in production processes in order to report product quality to the management. The audits are based on defects, defectives or demerits. Each sampled product is inspected, and the defects are assessed whenever the product fails to meet engineering requirements. The results are then compared to a quality standard, a target value reflecting a tradeoff between manufacturing cost, operating costs and customer needs. For audits based on defects or defectives, the standards are expressed in terms of defects or defectives per unit. For audits based on demerits, the standards are derived from fundamental defect per unit of count of A, B, C, D type defects (see Hoadley, 1981).

The Quality Measurement Plan (QMP), developed by Hoadley (1981), is a statistical method for analyzing discrete quality audit data which consist of the expected number of defects given standard quality. This plan was implemented for reporting the quality assurance audit results to Bell system management starting with the seventh period of 1980. The QMP is based on an empirical Bayes model of the audit-sampling process. It uses the past sample indices, but makes inference about the current quality. The method represented a considerable improvement in the statistical power for detecting substandard quality as compared with the old rules based on the T-rate system, evolved from the work of Shewart, Dodge and others, starting in the 1920s.

In spite of its wide publicity, QMP has been criticized on several grounds (see for example Barlow and Irony, 1992). The main criticism is that Hoadley's original procedure is at best heuristic, and a full Bayesian implementation of the procedure will require high-dimensional numerical integration. Hoadley's original procedure involves a Poisson likelihood with a gamma prior with the parameters of the gamma prior estimated from the marginal likelihood (after integrating with respect to the parameters of interest). However, the empirical Bayes versions of posterior means and variances as given by Hoadley (1981) are based on the assumption of independence of certain variables which usually fails to hold, especially for small samples. These points will be made specific in Section 3.3.

The primary objective of this chapter is to provide a hierarchical Bayes (HB) refinement of Hoadley's QMP. Such a HB procedure will avoid the ad hoc approximations needed in Hoadley's solution. Second, the present method will provide yet another illustration of the powerful Markov chain Monte Carlo integration technique which is gaining rapid popularity in recent years.

The outline of the remaining sections is as follows. In Section 3.2, we provide the notations and assumptions needed to describe the QMP model. In Section 3.3,

we describe the HB model and contrast it with Hoadley's (1981) model. Based on the present hierarchical model, we have found the posterior distributions of the parameters of interest as well as the posterior means and variances by using the Gibbs sampling technique (Gelfand and Smith, 1990; Gelfand et al., 1990). The Gibbs sampling method requires generating samples from different posterior distributions. In our derivation, one of the posterior distributions is known only up to a multiplicative constant. Accordingly, an accept-reject algorithm is used to generate samples from such a posterior. However, since this posterior turns out to be log-concave, we have been able to use the adaptive rejection sampling algorithm of Gilks and Wild (1992). A similar application of adaptive rejection sampling appears in George, Makov and Smith (1993). Besides giving the formulas for posterior means and variances, we have provided in this section a brief description of the Gibbs sampler as well as a description of the adaptive rejection sampling scheme. Finally, as a possible approximate solution, we have also discussed the Laplace approximation method (see Tierney and Kadane, 1986; Kass, Tierney and Kadane, 1989; Kass and Steffey, 1989) in the present context.

Section 3.4 contains the actual analysis of the data, We have provided the Bayes estimates and the associated standard errors of the current quality index using the HB model introduced in Section 3.3. We have also shown that the Laplace method may provide a poor approximation in this situation due to a heavily skewed posterior density.

Irony et al. (1992) have recently used an additive model and a multiplicative model as alternatives to QMP. The additive model deals with production processes that degrade as time goes by (processes that age for instance). The multiplicative model is appropriate for processes that improve with time (e.g. processes that depend on learning). In contrast, the QMP model of Hoadley assumes that the process average, say θ, although unknown is fixed. In reality, however, θ may be changing. In order

to handle this, the QMP procedure uses a moving window of six periods of data to infer on the current quality index. This underscores the importance of small sample inference associated with QMP procedures.

3.2 Notations and Assumptions

Suppose there are T rating periods: t=1,...,T, where T is the current period. For period t, we have the following data from the audit:

n_t = audit sample size;

x_t = number of defects in the audit sample;

s = standard number of defects per unit;

$e_t = sn_t$ = expected number of defects in the audit sample when the quality standard is met;

$I_t = x_t/e_t$ = defect index of the current sample.

ASSUMPTIONS : $x_t \sim \text{Poisson}(n_t\, \lambda_t)$ where λ_t is the defect rate per unit. Reparameterize λ_t as $\theta_t = \lambda_t/s$ = quality index at rating period t. Then $\theta_t = 1$ is the standard value and also $x_t \mid \theta_t \sim \text{Poisson}(e_t\theta_t)$.

The parameter of interest is θ_T, the current quality index. The objective is to derive the posterior distribution of θ_T given the data $\boldsymbol{x}$, which include the past data $(x_1, \cdots, x_{T-1})$ and the current data x_T.

3.3 Hierarchical Bayes Model

The following hierarchical Bayes (HB) model is considered:

I. Conditional on $\theta_1, \cdots, \theta_T$, α and β, $x_t \overset{ind}{\sim} \text{Poisson}(e_t\theta_t)$;

II. Conditional on α and β, $\theta_t \overset{\text{i.i.d}}{\sim} \text{Gamma}(\alpha, \beta)$, where a Gamma$(\alpha, \beta)$ variable, say Z has pdf $f(z \mid \alpha, \beta) = \exp(-\alpha z)\, z^{\beta-1}\, \alpha^{\beta} / \Gamma(\beta)$, $z > 0$, $\alpha > 0$, $\beta > 0$;

III. Marginally α and β have joint pdf

$$\pi(\alpha,\beta) \propto \alpha^{-1}\beta^{-a} \ (a > 1).$$

While doing the data analysis in Section 3.4, we shall consider several choices of a.

Jeffreys' prior is most widely used as a noninformative prior. This prior is proportional to the positive square root of the determinant of the expected Fisher-information matrix. For the Gamma(α,β) density, the expected Fisher-information matrix is given by

$$I(\alpha,\beta) = \begin{pmatrix} E\left(-\frac{\partial^2 log\, g}{\partial\alpha^2}\right) & E\left(-\frac{\partial^2 log\, g}{\partial\alpha\partial\beta}\right) \\ E\left(-\frac{\partial^2 log\, g}{\partial\alpha\partial\beta}\right) & E\left(-\frac{\partial^2 log\, g}{\partial\beta^2}\right) \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\alpha^2} & \frac{-1}{\alpha} \\ \frac{-1}{\alpha} & \frac{\partial^2 log\Gamma(\beta)}{\partial\beta^2} \end{pmatrix}.$$

Hence, Jeffreys' prior for (α,β) is given by

$$\pi(\alpha,\beta) \propto |I(\alpha,\beta)|^{1/2} = \frac{1}{\alpha}\left[\beta\,\frac{\partial^2 log\Gamma(\beta)}{\partial\beta^2} - 1\right]^{1/2}. \tag{3.3.1}$$

This prior has limited practical utility due to appearance of the complicated trigamma function. However, using Stirling's approximation

$$\Gamma(\beta) \approx \sqrt{2\pi}\, e^{-\beta}\, \beta^{\beta-1/2}.$$

Thus,

$$\frac{\partial^2 log\Gamma(\beta)}{\partial\beta^2} \approx \frac{1}{2\beta^2} + \frac{1}{\beta}. \tag{3.3.2}$$

Substitution of (3.3.2) into (3.3.1) yields

$$\pi(\alpha,\beta) \propto \frac{1}{\alpha}\left[\frac{1}{2\beta} + 1 - 1\right]^{1/2} \propto \alpha^{-1}\beta^{-1/2}. \tag{3.3.3}$$

However, this leads to an improper posterior for θ_i's. To avoid this, we take a prior of the form given in III.

The present hierarchical model is closely akin to a similar model of George, Makov and Smith (1993). The difference occurs at the third stage of the hierarchical model where George et al. use proper independent gamma priors for α and β, whereas we are using some diffuse gamma priors instead. We prefer to use the present class of priors for this problem due to lack of prior elicitation, making subjective analysis more difficult to justify.

Based on the present hierarchical model, a subjective Bayesian approach to find the posterior distribution of θ_T given $\boldsymbol{x}$ proceeds as follows

$$(i)\ \theta_T \mid \boldsymbol{x},\ \alpha,\ \beta\ \sim\ \mathrm{Gamma}(\mathrm{e_T} + \alpha, \mathrm{x_T} + \beta); \tag{3.3.4}$$

$$(ii)\ p(\alpha,\ \beta \mid \boldsymbol{x})\ \propto\ \beta^{-a}\alpha^{T\beta-1}\prod_{t=1}^{T}(\alpha + e_T)^{(x_T+\beta)}\ \prod_{t=1}^{T}\{\Gamma(x_t+\beta)/\Gamma(\beta)\} \tag{3.3.5}$$

Lemma 3.1 Suppose $x_t \geq 1$ for all $t = 1, \cdots, T$. Then $\int_0^\infty\int_0^\infty\ p(\alpha,\ \beta \mid \boldsymbol{x})\,d\alpha d\beta < \infty$ provided $\sum_{t=1}^{T} x_t\ \geq\ T\ >\ a\ >\ 1.$

Proof of Lemma 3.1 Note $p(\alpha,\ \beta \mid \boldsymbol{x})\ \propto\ h(\alpha, \beta)$, where

$$h(\alpha, \beta)\ =\ \beta^{-a}\alpha^{T\beta-1}\prod_{t=1}^{T}(\alpha + e_t)^{(x_t+\beta)}\ \prod_{t=1}^{T}\{\Gamma(x_t+\beta)/\Gamma(\beta)\}.$$

In what follows, we shall use the notation $K\ (>0)$ for a generic constant which may depend on $\boldsymbol{x}$, but not on α and β.

First, using Stirling's bounds for factorials for $\beta\ \geq\ \beta_0\ >\ 0$,

$$\prod_{t=1}^{T}[\Gamma(\beta + x_t)/\Gamma(\beta)]\ =\ \prod_{t=1}^{T}\left[\frac{\beta}{\beta + x_t}\ \frac{\Gamma(\beta + x_t + 1)}{\Gamma(\beta+1)}\right]$$

$$\leq\ \prod_{t=1}^{T}\left\{\frac{e^{-(\beta+x_t)}(\beta + x_t)^{\beta+x_t+\frac{1}{2}}e^{\frac{1}{12(\beta+x_t)}}}{e^{-\beta}\ \beta^{\beta+\frac{1}{2}}}\right\}$$

$$\leq K\beta^{\sum x_t} \prod_{t=1}^{T}(1 + \frac{x_t}{\beta})^{\beta + x_t + \frac{1}{2}}$$

$$\leq K\beta^{\sum x_t} exp\left[\sum x_t(x_t + \frac{1}{2})/\beta_0\right] \leq K\beta^{\sum x_t}$$

(since $\sum_{t=1}^{T}(\beta + x_t + \frac{1}{2})log\,(1 + \frac{x_t}{\beta})$

$$\leq \sum_{t=1}^{T}\left(\beta + x_t + \frac{1}{2}\right)\frac{x_t}{\beta} = \sum x_t + \frac{\sum x_t(x_t + \frac{1}{2})}{\beta} \leq \sum x_t + \frac{\sum x_t(x_t + \frac{1}{2})}{\beta_0} \;).$$

For $0 < \beta < \beta_0$,

$$\begin{aligned} \Gamma(\beta + x_t + 1)/\Gamma(\beta + 1) &= \int_0^\infty e^{-z} z^{\beta + x_t}\, dz \;\Big/ \int_0^\infty e^{-z} z^{\beta}\, dz \\ &= E_\beta(z^{x_t}), \end{aligned}$$

where $Z \sim \text{Gamma}(1, \beta + 1)$. Using the MLR property of gamma distributions $E_\beta(z^{x_t}) \uparrow$ in β so that $E_\beta(z^{x_t}) \leq E_{\beta_0}(z^{x_t})$. Now,

$$\prod_{t=1}^{T}[\Gamma(\beta + x_t)/\Gamma(\beta)] \leq \beta^T(\prod_{t=1}^{T} x_t^{-1})\prod_{t=1}^{T} E_{\beta_0}(z^{x_t}) \leq K\beta^T.$$

Hence, writing $c = \sum_{t=1}^{T}[log\,(\alpha + e_t) - log\,\alpha]$, it follows that

$$\begin{aligned} \int_0^\infty h(\alpha, \beta)\, d\beta &= \left(\int_0^{\beta_0} + \int_{\beta_0}^\infty\right)\alpha^{-1}\prod_{t=1}^{T}(\alpha + e_t)^{-x_t}\, exp(-c\beta)\beta^{-a}\prod_{t=1}^{T}\frac{\Gamma(\beta + x_t)}{\Gamma(\beta)}\, d\beta \\ &\leq K\alpha^{-1}\prod_{t=1}^{T}(\alpha + e_t)^{-x_t}\left[\int_0^{\beta_0} exp(-c\beta)\beta^{T-a}\, d\beta \right. \\ &\qquad \left. + \int_{\beta_0}^\infty exp(-c\beta)\beta^{\sum x_t - a}\, d\beta\right] \\ &\leq K\alpha^{-1}\prod_{t=1}^{T}(\alpha + e_t)^{-x_t}\left[c^{-T+a-1}\,\Gamma(T - a + 1)\right. \\ &\qquad \left. + c^{-\sum x_t + a - 1}\Gamma(\sum x_t - a + 1)\right] \end{aligned}$$

$$\leq \; K\alpha^{-1}\prod_{t=1}^{T}(\alpha+e_t)^{-x_t}\left(c^{-T+a-1} \; + \; c^{-\sum x_t+a-1}\right)$$

$$\leq \; K\alpha^{-1}(\alpha+e_{min})^{-\sum x_t}\left(\{log\,(\alpha+e_{min})-log\,\alpha\}^{-T+a-1}\right.$$

$$\left. + \; \{log\,(\alpha+e_{min})-log\,\alpha\}^{-\sum x_t+a-1}\right) \qquad (3.3.6)$$

(since $c = \sum_{t=1}^{T}(log\,(\alpha+e_t)-log\,\alpha) \geq T(log\,(\alpha+e_{min})-log\,\alpha)$, where $e_{min} = \min_{1\leq t\leq T} e_t$, and $\sum x_t > a$ and $T > a$).

Consider an interior point d of $(0,\infty)$ where $d > \frac{1}{2}e_{min}$. Then,

$$\int_d^\infty \alpha^{-1}(\alpha+e_{min})^{-\sum x_t}\,[log\,(\alpha+e_{min})-log\,\alpha]^{-T+a-1}\;d\alpha$$

$$= \int_d^\infty \alpha^{-1}\alpha^{-\sum x_t}\left[log\,(1+\frac{e_{min}}{\alpha})\right]^{-T+a-1}\;d\alpha$$

$$\leq \int_d^\infty \alpha^{-1}\alpha^{-\sum x_t}\left(\frac{e_{min}}{\alpha}-\frac{e_{min}^2}{2\alpha^2}\right)^{-T+a-1}\;d\alpha$$

$$\leq \int_d^\infty \alpha^{-1-\sum x_t}\left(\frac{e_{min}}{\alpha}\right)^{-T+a-1}\left(1-\frac{e_{min}}{2d}\right)^{-T+a-1}\;d\alpha$$

$$\leq K\int_d^\infty \alpha^{T-\sum x_t-a}\,d\alpha \;\leq\; K\int_d^\infty \alpha^{-a}\,d\alpha \;<\; \infty \qquad (3.3.7)$$

(since $\sum_t x_t \geq T$ and $a > 1$).

Similar calculations yield

$$\int_d^\infty \alpha^{-1}(\alpha+e_{min})^{-\sum x_t}\;[log\,(\alpha+e_{min})-log\,\alpha]^{-\sum x_t+a-1}\;d\alpha$$

$$\leq \; K\int_d^\infty \alpha^{\sum x_t-\sum x_t-a}\,d\alpha \;<\; \infty. \qquad (3.3.8)$$

Now, observe that

$$\int_0^d \alpha^{-1}(\alpha+e_{min})^{-\sum x_t}\;[log\,(\alpha+e_{min})-log\,\alpha]^{-T+a-1}\;d\alpha$$

$$= \int_0^d (\alpha + e_{min})^{-\sum x_t + 1} \alpha^{-1} (\alpha + e_{min})^{-1} \; [log\,(\alpha + e_{min}) - log\,\alpha]^{-T+a-1} \; d\alpha$$

$$\leq (e_{min})^{-\sum x_t + 1} \left[\frac{(log\,(\alpha + e_{min}) - log\,\alpha)^{-T+a}}{(-e_{min})(-T + a)} \right]_{\alpha=0}^{\alpha=d}$$

$$< \infty \quad (\text{since T} > \text{a}). \tag{3.3.9}$$

Again,

$$\int_0^d \alpha^{-1} (\alpha + e_{min})^{-\sum x_t} \; [log\,(\alpha + e_{min}) - log\,\alpha]^{-\sum x_t + a - 1} \; d\alpha$$

$$\leq (e_{min})^{-\sum x_t + 1} \left[\frac{(log\,(\alpha + e_{min}) - log\,\alpha)^{-\sum x_t + a}}{(-e_{min})(-\sum x_t + a)} \right]_{\alpha=0}^{\alpha=d}$$

$$< \infty \quad (\text{since } \textstyle\sum \text{x}_t > \text{a}) \tag{3.3.10}$$

Combine (3.3.6)-(3.3.10) to get $\int_0^\infty \int_0^\infty h(\alpha, \beta)\, d\alpha\, d\beta < \infty$.

The above result ensures (due to (3.3.4) and (3.3.5)) that the posterior pdf of θ_T given $\boldsymbol{x}$ is a proper pdf under the same condition. If this is the case, then using (3.3.4), the posterior mean and the posterior variance are given by

$$E\,[\theta_t \mid \boldsymbol{x}] \;=\; E\,[E\,(\theta_t \mid \alpha,\, \beta,\, \boldsymbol{x}) \mid \boldsymbol{x}] \;=\; E\left[(x_t + \beta)(e_t + \alpha)^{-1} \mid \boldsymbol{x}\right] \tag{3.3.11}$$

$$\begin{aligned} V\,[\theta_t \mid \boldsymbol{x}] &= E\,[V\,(\theta_t \mid \boldsymbol{x}, \alpha, \beta) \mid \boldsymbol{x}] \;+ V\,[E\,(\theta_t \mid \boldsymbol{x}, \alpha, \beta) \mid \boldsymbol{x}] \\ &= E\,[(x_t + \beta)(e_t + \alpha)^{-2} \mid \boldsymbol{x}] \;+ V\left[(x_t + \beta)(e_t + \alpha)^{-1} \mid \boldsymbol{x}\right] \end{aligned} \tag{3.3.12}$$

Hoadley (1981) uses the notation $\theta = \beta/\alpha$, the prior mean for the θ_t's which he calls "process average" and $\gamma = \beta/\alpha^2$, the prior variance, which he calls "process variance". Writing $w_T = \alpha/(e_T + \alpha)$, it follows from (3.3.11) that

$$E\,[\theta_t \mid \boldsymbol{x}] \;=\; E\,[(1 - w_T) I_T \;+\; w_T \theta \mid \boldsymbol{x}]. \tag{3.3.13}$$

If the prior parameters α and β were known, as in subjective Bayes analysis, then the posterior mean would be $(1 - w_T) I_T + w_T \theta$, a weighted average of the current defect

index I_T and the process average θ. If α is small compared to e_T, i.e. the sample evidence outweighs the prior evidence, then the weighted average leans more toward I_T, the current index. The opposite is the case when α is large compared to e_T.

Hoadley starts with (3.3.13), but unlike our stage III, does not assume any hyperprior for α and β. Instead he estimates α and β from the marginal distribution of $\boldsymbol{x}$ after integrating out $\theta_1, \cdots, \theta_T$. In this way an approximation $(1 - \hat{w}_T)I_T + \hat{w}_T\hat{\theta}$ for the expression in the right hand side of (3.3.13) can be made where $\hat{w}_T$ and $\hat{\theta}$ are EB estimators of w_T and θ respectively. But Hoadley (see his p 233) seems to argue $E(w_T\theta \mid \boldsymbol{x}) \approx E(w_T \mid \boldsymbol{x})\, E(\theta \mid \boldsymbol{x})$ and then approximate each of $E(w_T \mid \boldsymbol{x})$ and $E(\theta \mid \boldsymbol{x})$. The posterior uncorrelation of w_T and θ does not hold in general. Next note that

$$
\begin{aligned}
&V\left[(x_t + \beta)(e_t + \alpha)^{-1} \mid \boldsymbol{x}\right] = \\
&\quad = V\left[(1 - w_T)I_T + w_T\theta \mid \boldsymbol{x}\right] = V\left[w_T(\theta - I_T) \mid \boldsymbol{x}\right] \\
&\quad = V\left[w_T(\theta - \hat{\theta}_B + \hat{\theta}_B - I_T) \mid \boldsymbol{x}\right], \qquad (3.3.14)
\end{aligned}
$$

where $\hat{\theta}_B = E(\theta \mid \boldsymbol{x})$. Hoadley approximates the right hand side of (3.3.14) by $\hat{w}_T^2 V(\theta \mid \boldsymbol{x}) + (\hat{\theta} - I_T)^2 V(w_T \mid \boldsymbol{x})$. This approximation is much more questionable since neglecting $\mathrm{Cov}(w_T(\theta - \hat{\theta}_B), w_T(\hat{\theta}_B - I_T) \mid \boldsymbol{x})$ may be too much of a sacrifice. Second, the approximation of $V(w_T(\theta - \hat{\theta}_B) \mid \boldsymbol{x})$ by $\hat{w}_T^2 V(\theta \mid \boldsymbol{x})$ does not take into account the posterior dependence of w_T and θ.

The above does not undermine Hoadley's novel contribution. The main difficulty that he faced was that finding the posterior distribution of θ_T given $\boldsymbol{x}$ using the hierarchical Bayes model requires multidimensional integrals. The usual numerical integration tools are not very reliable in high dimensions. Monte Carlo numerical integration was not very popular in those days.

In the present study, we use Monte Carlo numerical integration to generate posterior distributions and associated means and variances. More specifically, we use Gibbs sampling originally introduced in Geman and Geman (1984), and more recently popularized by Gelfand and Smith(1990) and Gelfand et al. (1990). The method is described below.

Gibbs sampling is a Markovian updating scheme. Given an arbitrary starting set of values $U_1^{(0)}, \cdots, U_p^{(0)}$, we draw $U_1^{(1)} \sim [U_1 \mid U_2^{(0)}, \cdots, U_p^{(0)}]$, $U_2^{(1)} \sim [U_2 \mid U_1^{(1)}, U_3^{(0)}, \cdots, U_p^{(0)}], \cdots, U_p^{(1)} \sim [U_p \mid U_1^{(1)}, \cdots, U_{p-1}^{(1)}]$, where $[\cdot \mid \cdot]$ denotes the relevant conditional distributions. Thus, each variable is visited in the natural order and a cycle in this scheme requires p random variate generations. After k such iterations, one arrives at $(U_1^{(k)}, \cdots, U_p^{(k)})$. As k$\rightarrow \infty$, $(U_1^{(k)}, \cdots, U_p^{(k)}) \xrightarrow{d} (U_1, \cdots, U_p)$. Gibbs sampling through q replications of the aforementioned k - iterations generates q iid p-tuples $(U_{1j}^{(k)}, \cdots, U_{pj}^{(k)})$ (j =1,$\cdots$,q); $U_1, \cdots, U_p$ could possibly be vectors in the above scheme.

Using Gibbs sampling, the joint posterior pdf of $\boldsymbol{\theta}= (\theta_1, \cdots, \theta_T)$ is approximated by

$$q^{-1} \sum_{j=1}^{q} \left[\boldsymbol{\theta} \mid \boldsymbol{x}, \alpha = \alpha_j^{(k)}, \beta = \beta_j^{(k)}\right]. \tag{3.3.15}$$

The Gibbs sampling analysis is based on the following posterior distributions:

(i) $\theta_t \mid \boldsymbol{x}, \alpha, \beta \overset{ind}{\sim}$ Gamma$(e_t + \alpha\, ,\, x_t + \beta)$;

(ii) $\alpha \mid \boldsymbol{\theta}, \boldsymbol{x}, \beta \overset{ind}{\sim}$ Gamma$(\sum_{t=1}^{T} \theta_t\, ,\, T\beta)$;

(iii) $\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha$ has pdf p$(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha) \propto (\prod_{t=1}^{T} \theta_t)^{\beta-1} \beta^{-a} \frac{\alpha^{T\beta}}{(\Gamma(\beta))^T}$

To estimate the posterior moments, we use Rao-Blackwellized estimates as in Gelfand and Smith (1991). Using (3.3.11), $E(\theta_T \mid \boldsymbol{x})$ is approximated by

$$q^{-1} \sum_{j=1}^{q} \left(\frac{x_T + \beta_j^{(k)}}{e_T + \alpha_j^{(k)}} \right), \tag{3.3.16}$$

where as before, k denotes the number of iterations needed to generate a sample. Next using (3.3.12), $V(\theta_T \mid \boldsymbol{x})$ is approximated by

$$q^{-1}\sum_{j=1}^{q}\left(\frac{x_T+\beta_j^{(k)}}{(e_T+\alpha_j^{(k)})^2}\right)+\left[q^{-1}\sum_{j=1}^{q}\left(\frac{x_T+\beta_j^{(k)}}{e_T+\alpha_j^{(k)}}\right)^2-\left(q^{-1}\sum_{j=1}^{q}\frac{x_T+\beta_j^{(k)}}{e_T+\alpha_j^{(k)}}\right)^2\right] \quad (3.3.17)$$

In implementing the Gibbs sampler, one should be able to draw samples from the conditional densities given in (i)-(iii). Simulation from the conditional densities (i) and (ii) which are both gamma densities can be done by standard methods. However, the posterior pdf of β given $\boldsymbol{\theta}$,$\boldsymbol{x}$ and α is known only up to a multiplicative constant. In order to simulate from this density, one general approach is to use the Metropolis-Hastings accept-reject algorithm.

Fortunately, the task becomes simpler for us because of the following result.

Lemma 3.2 *log* $p(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha)$ is a concave function of β if T > a.

Proof of Lemma 3.2 Consider $p(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha) \propto (\prod_{t=1}^{T}\theta_t)^{\beta-1}\beta^{-a}\frac{\alpha^{T\beta}}{(\Gamma(\beta))^T}$. Then

$$log\ p(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha) = C + (\beta-1)\sum_t log\ \theta_t - a\ log\ \beta + T\beta log\ \alpha - T log\ \Gamma(\beta)$$

where C is the norming constant. Hence,

$$\frac{\partial\ log\ p(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha)}{\partial\ \beta}$$

$$= \textstyle\sum_t log\ \theta_t + T\ log\ \alpha - \frac{a}{\beta} - T\frac{d}{d\beta}\ log\ \left[\frac{\Gamma(\beta+1)}{\beta}\right]$$

$$= \textstyle\sum_t log\ \theta_t + T\ log\ \alpha + (T-a)\frac{1}{\beta} - T\frac{\int_0^\infty e^{-z}\, z^\beta\, log\, z\, dz}{\int_0^\infty e^{-z}\, z^\beta\, dz} \quad (3.3.18)$$

Therefore,

$$\frac{\partial^2\ log\ p(\beta \mid \boldsymbol{\theta}, \boldsymbol{x}, \alpha)}{\partial\ \beta^2}$$

$$= -(T-a)\frac{1}{\beta^2} - T\left[\frac{\int_0^\infty e^{-z} z^\beta (\log z)^2\, dz}{\int_0^\infty e^{-z} z^\beta\, dz} - \left(\frac{\int_0^\infty e^{-z} z^\beta \log z\, dz}{\int_0^\infty e^{-z} z^\beta\, dz}\right)^2\right]$$

$$= -(T-a)\frac{1}{\beta^2} - T\, V_{\beta+1}(\log z) < 0 \qquad (3.3.19)$$

for T > a, where $z \sim$ Gamma(1, $(\beta+1)$)

Because of the log-concavity of this posterior density, we can use the adaptive rejection sampling algorithm of Gilks and Wild (1992) to simulate from this density.

The adaptive rejection sampling is a black box technique for sampling from any univariate log-concave probability density function $f(x)$. The algorithm is based on the fact that any concave function can be bounded by a rejection envelope and a squeezing function which are piecewise exponential functions, constructed by tangents at, and the squeezing function by chords between, evaluated sampled points on the function over its domain. As sampling proceeds, the rejection envelope and the squeezing function converge to the density function, and hence the method is adaptive.

We now describe the adaptive rejection sampling in a general framework. Let $f(x)$ be a probability density function with domain D. It is assumed that D is connected and $f(x)$ is continuous and differentiable everywhere in D and that $h(x) = \ln f(x)$ is concave everywhere in D. Consider m abscissa points in D: $x_1 \le x_2 \le \cdots \le x_m$. Let $T_m = \{x_i;\ i = 1, \cdots, m\}$. For $j = 1, \cdots, m-1$ the tangents to $h(x)$ at x_j and x_{j+1} in T_m intersect at

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}\,. \qquad (3.3.20)$$

The rejection envelope on T_m is defined as $\exp(u_m(x))$ where $u_m(x)$ is a piecewise linear upper hull of the form

$$u_m(x) = h(x_j) + (x - x_j)h'(x_j) \text{ for } x \in [z_{j-1}, z_j],\ j = 1, \cdots, k \qquad (3.3.21)$$

where z_0 is the lower bound of D(or $-\infty$ if D is unbounded below) and z_m is the upper bound of D(or $+\infty$ if D is unbounded above). The squeezing function on T_m is defined as $\exp(l_m(x))$, where $l_m(x)$ is a piecewise linear lower hull formed from the chords between adjacent abscissae in T_m and is of the form

$$l_m(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j} \tag{3.3.22}$$

for $j = 1, \cdots, m-1$. For $x < x_1$ or $x > x_m$ $l_m(x) = -\infty$. Also, define the following function

$$s_m(x) = exp(u_m(x)) \Big/ \int_D exp(u_m(x'))\, dx' \,. \tag{3.3.23}$$

The concavity of $h(x)$ ensures that $l_m(x) \leq h(x) \leq u_m(x)$ for all x in D. To sample n points independently from $f(x)$ the following steps are performed: (1) *Initialization step*, (2) *Sampling step* and (3) *Updating step*.

Initialization step : Intialize the abscissa points in T_m. If D is unbounded, below then x_1 is chosen such that $h'(x_1) > 0$ and if D is unbounded above, then x_m is chosen such that $h'(x_m) < 0$. The functions $u_m(x)$, $l_m(x)$ and $s_m(x)$ are found from equations (3.3.21),(3.3.22) and (3.3.23) respectively.

Sampling step : x^* is sampled from $s_m(x)$ and a value w is sampled independently from Uniform(0,1) distribution. The squeezing test is performed as follows: if

$$w \leq exp\{l_m(x^*) - u_m(x^*)\}$$

then x^* is accepted. Else $h(x^*)$ and $h'(x^*)$ are evaluated and the rejection test is performed: if

$$w \leq exp\{h(x^*) - u_m(x^*)\}$$

then x^* is accepted; otherwise x^* is rejected.

Updating step : If $h(x^*)$ and $h'(x^*)$ were evaluated at the *sampling step* then x^* was included in T_m to form T_{m+1} and the elements of T_{m+1} were relabelled in

ascending order. Then the functions $u_{m+1}(x)$, $l_{m+1}(x)$ and $s_{m+1}(x)$ from equations (3.3.21),(3.3.22) and (3.3.23) respectively are evaluated. We return to the *sampling step* if n points have not been accepted.

As an alternative to Gibbs sampling the posterior moments of θ_T given $\boldsymbol{x}$ can also be obtained using the Laplace method of approximation (see Tierney and Kadane, 1986). Note that

$$\begin{aligned} E\left[\theta_T \mid \boldsymbol{x}\right] &= E\left[\left(\frac{x_T+\beta}{e_T+\alpha}\right) \mid \boldsymbol{x}\right] \\ &= \left(\frac{\int\int \left(\frac{x_t+\beta}{e_t+\alpha}\right)\, p(\alpha,\, \beta,\mid\, \boldsymbol{x})\, d\alpha\, d\beta}{\int\int\, p(\alpha,\, \beta,\mid\, \boldsymbol{x})\, d\alpha\, d\beta}\right) \end{aligned} \tag{3.3.24}$$

Setting

$$\begin{aligned} L &= \, log\, p(\alpha,\, \beta,\mid\, \boldsymbol{x}) \\ &= \, log\, C + (T\beta - 1) log\, \alpha - a\, log\, \beta + \textstyle\sum_t\, log\, \Gamma(x_t+\beta) - T\, log\, \Gamma(\beta) \\ &\qquad - \textstyle\sum_t (x_t+\beta) log\, (e_t+\alpha)\ , \end{aligned} \tag{3.3.25}$$

where C is a norming constant, and

$$L^* \;=\; log\left(\frac{x_T+\beta}{e_T+\alpha}\right) + log\, p(\alpha,\, \beta,\mid\, \boldsymbol{x}) \tag{3.3.26}$$

produces the approximation

$$\hat{E}\left[\theta_t \mid \boldsymbol{x}\right] \;=\; \left(\frac{det\, \Sigma^*}{det\, \Sigma}\right)^{1/2} exp\left\{L^*(\hat{\alpha}^*, \hat{\beta}^*) \;-\; L(\hat{\alpha}, \hat{\beta})\right\} \tag{3.3.27}$$

to $E\left[\theta_t \mid \boldsymbol{x}\right]$, where $(\hat{\alpha}^*, \hat{\beta}^*)$ and $(\hat{\alpha}, \hat{\beta})$ maximize L^* and L respectively and Σ^* and Σ are minus the inverse Hessians of L^* and L at $(\hat{\alpha}^*, \hat{\beta}^*)$ and $(\hat{\alpha}, \hat{\beta})$ respectively. The approximation (3.3.27) is referred to as the first order Laplace approximation. A

similar approximation applies to the posterior variance when one writes

$$V\left[\frac{x_t+\beta}{e_t+\alpha}\mid \boldsymbol{x}\right] = E\left[\left(\frac{x_t+\beta}{e_t+\alpha}\right)^2 \mid \boldsymbol{x}\right] - \left\{E\left[\left(\frac{x_t+\beta}{e_t+\alpha}\right) \mid \boldsymbol{x}\right]\right\}^2 , \qquad (3.3.28)$$

substitutes (3.3.28) in (3.3.12), and uses calculation similar to (3.3.25) and (3.3.26) for each term in (3.3.12) to arrive at an expression similar to (3.3.27).

It should be noted in the present context that since $(x_T+\beta)/(e_T+\alpha) > 0$, second order Laplace approximation is automatically achieved using the present approach as in Tierney and Kadane (1986). One does not need to appeal to Kass and Steffey (1989) which provides second order Laplace approximation even when the integrand is not necessarily positive.

3.4 An Example

The example in this section considers the same defect data as given in Hoadley (1981). Hoadley's primary goal was to compare QMP with then existing T-rate method. Our objective is to compare and contrast the present HB method with Hoadley's EB method.

In deriving the HB estimates of the present chapter, we have considered Gibbs sampler with a burn-in sample of 2000, subsequent iterations being 50 to get one sample. A sample of size 10,000 is taken to obtain the Monte-Carlo estimates, as stability seems to be achieved with this sample size.

Table 3.1 provides the expressions for e_t (expected number of defects in the audit sample when the quality standard is met for period t), I_t (defect index of the current sample for period t), $\hat{\theta}_t$ (posterior mean of quality index at period t using Hoadley's QMP), V_t (posterior variance of quality index at period t using Hoadley's QMP), $\hat{\theta}^{HB}_{t(a)}$ (posterior mean of quality index at period t using the present HB method for different choices $a = 2, 3$ and 4), $\hat{\sigma}^{HB}_{t(a)}$ (posterior variance of quality index at period t

using the present HB method) once again for $a = 2, 3$ and 4, $\hat{\theta}^L_{t(a)}$ (posterior mean of quality index at period t using Laplace's approximation), $\hat{\sigma}^L_{t(a)}$ (posterior variance of quality index at period t using Laplace's approximation) for $a = 2, 3, 4$. These figures are provided for the same 9 time periods $t = 1, \cdots, 9$ as given in Table 4 of Hoadley (1981).

An inspection of Table 3.1 reveals that there can be significant difference in the estimates of θ_t as given in Hoadley and the HB estimates given in this chapter. Also, the HB procedure is somewhat sensitive to the choice of "a" as different choices of "a" can lead to slightly different point estimates. The posterior variances obtained by the HB approach also are dissimilar to the approximate expressions given by Hoadley. They are substantially smaller for periods 3 and 4, but on the other hand, are much bigger for periods 1,2,5,6,7,8 and 9. Also, there is some sensitivity of the proposed HB method regarding the choice of "a". However, these differences are not as drastic as compared to the differences of the HB method and Hoadley's approximations.

Laplace's approximations do not work very well in the present context. The main reason is that the joint posterior distribution of (α, β) given the data is highly skewed. It is a folklore that the Laplace approximation works well only when the posterior distribution is close to Gaussian. The bigger the departure from normality, greater is the inadequacy of Laplace's method. The present example provides yet another illustration of this phenomenon. Situation may improve if one works with some transformation of α and β. This idea is not explored in this chapter, and will be a topic for future study.

TABLE 3.1: SUMMARY DATA FOR QMP

Periods	I_t	e_t	$\hat{\theta}_t$	V_t	$\hat{\theta}_{t(a)}$						$\hat{\sigma}_{t(a)}$					
					$\hat{\theta}^{HB}_{t(2)}$	$\hat{\theta}^{L}_{t(2)}$	$\hat{\theta}^{HB}_{t(3)}$	$\hat{\theta}^{L}_{t(3)}$	$\hat{\theta}^{HB}_{t(4)}$	$\hat{\theta}^{L}_{t(4)}$	$\hat{\sigma}^{HB}_{t(2)}$	$\hat{\sigma}^{L}_{t(2)}$	$\hat{\sigma}^{HB}_{t(3)}$	$\hat{\sigma}^{L}_{t(3)}$	$\hat{\sigma}^{HB}_{t(4)}$	$\hat{\sigma}^{L}_{t(4)}$
1	2.74	2.78	1.81	0.36	2.52	2.70	2.65	2.75	2.71	2.76	0.78	0.90	0.87	0.94	0.93	0.96
2	3.52	3.60	2.65	0.56	3.01	3.35	3.27	3.45	3.40	3.49	0.83	0.91	0.90	0.94	0.93	0.96
3	8.00	0.29	4.19	6.97	3.30	3.98	4.23	4.95	5.15	5.71	1.95	2.77	2.72	3.39	3.38	3.79
4	6.81	0.23	3.83	8.47	2.99	3.43	3.71	4.19	4.43	4.79	1.81	2.58	2.57	3.24	3.27	3.65
5	1.57	7.19	1.79	0.15	1.73	1.67	1.67	1.64	1.63	1.62	0.44	0.46	0.46	0.47	0.46	0.47
6	1.61	7.19	1.57	0.14	1.76	1.71	1.71	1.67	1.67	1.65	0.44	0.47	0.46	0.47	0.47	0.47
7	1.50	1.32	1.38	0.091	1.69	1.62	1.62	1.57	1.57	1.55	0.45	0.47	0.46	0.47	0.47	0.47
8	1.04	4.92	1.26	0.11	1.42	1.27	1.27	1.18	1.18	1.13	0.48	0.49	0.48	0.48	0.48	0.47
9	3.10	1.40	2.08	1.02	2.62	2.91	2.84	3.03	2.97	3.09	1.01	1.24	1.19	1.34	1.31	1.40

CHAPTER 4

BAYESIAN ANALYSIS OF CATEGORICAL SURVEY DATA

4.1 Introduction

Small area estimation is gaining increasing importance in survey sampling due to growing demand of reliable small area statistics both from public and private sectors. In typical small area estimation problems, there exist a large number of local areas, but samples available from an individual area are not usually adequate to achieve accuracy at a specified level. The reason behind this is that the original survey was designed to provide specific accuracy at a much higher level of aggregation than that for local areas. This makes it a necessity to "borrow strength" or connect these local areas explicitly or implicitly through models. In consequence, an estimate for a particular local area utilizes information from similar neighbouring areas. For an early history as well the recent developments on small area estimation, the reader is referred to the survey article of Ghosh and Rao (1991).

For quite some time now, Bayesian methods have been applied very extensively for solving small area estimation problems. Particularly effective in this regard has been the hierarchical or empirical Bayes (HB or EB) approach which are especially suited for a systematic connection of the local areas through models. For the general discussion of the EB or HB methodology in the small area estimation context, the reader is referred to Ghosh and Meeden (1986), Ghosh and Lahiri (1987), Datta and Ghosh (1990) among others.

However, the development to date has mainly concentrated on numerical valued variates. Often the survey data are categorical, for which the HB or EB analysis suitable for continuous variates is not very appropriate. It is only recently that some work has started on the analysis of binary survey data. McGibbon and Tomberlin (1989) obtain small area estimates of proportions via EB techniques, while Stasny (1991) uses a HB model to estimate the probability that an individual has a certain characteristic. She uses data from the national crime survey (NCS) to estimate the probability of being victimized, and the data from the current population survey (CPS) to estimate the probability of being unemployed. Stroud (1991) develops a general HB methodology for binary data, and subsequently (Stroud, 1992) provides a comprehensive treatment of binary categorical survey data encompassing simple random, stratified, cluster and two-stage sampling as well as two-stage sampling within strata.

However, often the survey data, by nature, are more appropriately classified into several categories instead of two. Simple examples of such multi-category responses are choice of transportation to take to work (drive, bus, subway, walk, bicycle), concensus on an opinion (strongly agree, agree, disagree, and strongly disagree), political ideology (liberal, moderate, and conservative). To our knowledge, hardly any EB or HB analysis seems available for such data. The objective of this chapter is to provide a general HB methodology related to inference for data on items containing three or more possible responses. The analysis is done within the framework of two-stage sampling within strata. As a specific example to be considered in this chapter, we cite the recent Canada Youth and AIDS study (King et al., 1988). In the different provinces of Canada, children within selected schools were asked the question "how often have you had sexual intercourse?" There were four response categories: never, once, a few times, and often. Stroud (1991) analyzed the data by collapsing the four categories into two : "often" and "not often"; but a more elaborate analysis involving

all the four categories is bound to be more informative. We shall present a general HB methodology which will be found adequate to handle data of the above type.

The outline of the remaining sections is as follows. In Section 4.2, we present a general HB algorithm for inference based on generalized linear models. The inference includes but is not restricted to the important logistic regression or log-linear models. As mentioned earlier, the method is described when there is two-stage sampling within the strata, and in this way, our method extends the work of Zeger and Karim (1991) who consider one stage sampling within strata. As in Zeger and Karim (1991), we find the necessary posterior distributions by using the Gibbs sampling (see Geman and Geman, 1984; Gelfand and Smith, 1990), but there is one crucial simplification in this chapter. We have identified the log-concavity of several densities which are known only up to a multiplicative constant, and in this way have been able to use the general adaptive rejection sampling of Gilks and Wild (1992) in contrast to the more complex direct Metropolis-Hastings algorithm as done in Zeger and Karim (1991). Also, in Section 4.2 of this chapter, we have contrasted the present method to that of Albert (1988) and of Leonard and Novick (1986).

Section 4.3 considers general multi-category survey data admitting a multinomial likelihood. However, we have viewed the multinomial distribution as the joint distribution of several independent Poisson variables conditional on their sum, and in this way, have been able to bring in directly the results of Section 4.2 for the analysis of multi-category survey data.

Finally, in Section 4.4, we have considered the Youth and AIDS data to illustrate the general methods described in Section 4.3.

4.2 Generalized Linear Models for Two-Stage Sampling Within Strata

Let Y_{ijk} denote the response (discrete or continuous) of the kth unit within the jth cluster in the ith stratum ($k = 1, \cdots, n_{ij}$; $j = 1, \cdots, c_i$; $i = 1, \cdots, m$). The

Y_{ijk} are assumed to be independent with pdf's

$$f(y_{ijk};\theta_{ijk},\phi_{ijk}) \;=\; exp\left[\phi_{ijk}^{-1}(y_{ijk}\theta_{ijk} - \psi_{ijk}) + \rho(y_{ijk};\phi_{ijk})\right] \tag{4.2.1}$$

Such a model is referred to as a generalized linear model (McCullagh and Nelder, 1989, p 28). The density (4.2.1) is parametrized with respect to the canonical parameters θ_{ijk} and scale parameters ϕ_{ijk}. It is assumed that the scale parameters ϕ_{ijk} are known.

The natural parameters θ_{ijk} are modelled as

$$\theta_{ijk} \;=\; \boldsymbol{x}_{ijk}^T\boldsymbol{\beta} \;+\; u_{ij} \;+\; \epsilon_{ijk}, \tag{4.2.2}$$

where the $\boldsymbol{x}_{ijk}$ (p x 1) are known design vectors, $\boldsymbol{\beta}$ (p x 1) is the unknown regression coefficient, u_{ij} are the random effects, and ϵ_{ijk} are the errors. It is assumed that u_{ij} and the ϵ_{ijk} are mutually independent with u_{ij} iid N(0, σ_B^2) and ϵ_{ijk} iid N(0,σ^2).

It is possible to represent (4.2.1) and (4.2.2) as a " conditionally independent" hierarchical model (see e.g. Kass and Steffey (1990)). Write $\lambda_{ij} \;=\; \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + u_{ij}$, $R_B \;=\; (\sigma_B^2)^{-1}$ and $R \;=\; (\sigma^2)^{-1}$. Then the hierarchical model is given by

I. Conditional on $\boldsymbol{\beta}$, $R_B = r_B$ and $R = r$, Y_{ijk} are mutually independent with a density of the form given in (4.2.1).

II. Conditional on $\boldsymbol{\beta}$, $R_B = r_B$ and $R = r$, $\theta_{ijk} \overset{ind}{\sim} \mathrm{N}(\lambda_{ij}, r^{-1})$.

III. Conditional on $\boldsymbol{\beta}$, $R_B = r_B$ and $R = r$, $\lambda_{ij} \overset{ind}{\sim} \mathrm{N}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}, r_B^{-1})$.

To complete the HB analysis, we assign the prior

IV. $\boldsymbol{\beta}$, R_B and R are mutually independent with $\boldsymbol{\beta} \sim$ uniform($\boldsymbol{R}^p$), $R_B \sim$ Gamma($\frac{1}{2}a$, $\frac{1}{2}b$) and $R \sim$ Gamma($\frac{1}{2}c$, $\frac{1}{2}d$).

[A rv Z is said to have a Gamma(α,β) distribution if it has a pdf of the form $f(z) \propto exp(-\alpha z)\, z^{\beta-1}\, I_{[z>0]}$, where $\alpha > 0$, $\beta > 0$]. We allow the possibility of diffuse gamma priors by allowing a, b, c and d to be zeroes or even negative.

We are interested in the posterior distribution of θ_{ijk} given the data y_{ijk} ($k \;=\; 1,\cdots,n_{ij}$; $j \;=\; 1,\cdots,c_i$; $i \;=\; 1,\cdots,m$). This is best accomplished by using the

Gibbs sampler (Geman and Geman (1984), Gelfand and Smith (1990)). For its implementation, the necessary posterior distributions based on (I)-(IV) are given as follows:

(i) $\boldsymbol{\beta} \mid \boldsymbol{\lambda}, \boldsymbol{y}, \boldsymbol{\theta}, r_B, r \sim \mathrm{N}\Big(\big(\sum_i \sum_j \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T\big)^{-1} \big(\sum_i \sum_j \lambda_{ij}\boldsymbol{x}_{ij}\big), r_B^{-1}\big(\sum_i \sum_j \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T\big)^{-1}\Big)$;

(ii) $\lambda_{ij} \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{\theta}, r_B, r \overset{ind}{\sim} \mathrm{N}\Big((r+r_B)^{-1}(r\sum_k \theta_{ijk} + r_B \boldsymbol{x}_{ij}^T\boldsymbol{\beta}), (r+r_B)^{-1}\Big)$;

(iii) $R \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{\theta}, r_B, \boldsymbol{\lambda} \sim \mathrm{Gamma}(\frac{1}{2}(c+\sum_i \sum_j \sum_k (\theta_{ijk}-\lambda_{ij})^2), \frac{1}{2}(d+\sum_{i=1}^m \sum_{j=1}^{c_i} n_{ij}))$;

(iv) $R_B \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{\theta}, r, \boldsymbol{\lambda} \sim \mathrm{Gamma}(\frac{1}{2}(a+\sum_i \sum_j (\lambda_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2), \frac{1}{2}(b+\sum_{i=1}^m c_i))$;

(v) $\theta_{ijk} \mid \boldsymbol{\beta}, \boldsymbol{y}, r_B, r, \lambda$ are mutually independent with

$$f(\theta_{ijk} \mid \boldsymbol{\beta}, \boldsymbol{y}, r_B, r, \boldsymbol{\lambda}) \;\propto\; \left\{exp\left(\theta_{ijk}y_{ijk} - \psi(\theta_{ijk}) - \frac{1}{2}r(\theta_{ijk}-\lambda_{ij})^2\right)\right\}$$

It is clear from the above that it is possible to generate samples from the normal and gamma distributions given in (i)-(iv). On the other hand, as evidenced in (v), the posterior distribution of θ_{ijk} given $\boldsymbol{\beta}$, $\boldsymbol{y}$, r_B, r and $\boldsymbol{\lambda}$ is known only up to a multiplicative constant, and accordingly one has to use a general accept-reject algorithm to generate samples from this pdf. Fortunately, the task becomes much simpler due to the following lemma establishing log-concavity of this posterior density, because then one can use the adaptive rejection sampling of Gilks and Wild (1992).

Lemma 4.1 *log* $f(\theta_{ijk} \mid \boldsymbol{\beta}, \boldsymbol{y}, r_B, r, \boldsymbol{\lambda})$ is a concave function of θ_{ijk}.

Proof of Lemma 4.1

$$\frac{\partial log\, f(\theta_{ijk} \mid \boldsymbol{\beta}, \boldsymbol{y}, r_B, r, \boldsymbol{\lambda})}{\partial \theta_{ijk}} = y_{ijk} - \psi'(\theta_{ijk}) - r(\theta_{ijk} - \lambda_{ij})$$

Hence,

$$\frac{\partial^2 log\, f(\theta_{ijk} \mid \boldsymbol{\beta}, \boldsymbol{y}, r_B, r, \boldsymbol{\lambda})}{\partial \theta_{ijk}^2} = -\psi''(\theta_{ijk}) - r < 0,$$

using the fact that $r > 0$ and $\psi''(\theta_{ijk}) = V(Y_{ijk} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, r, r_B, \boldsymbol{\lambda}) = V(Y_{ijk} \mid \boldsymbol{\theta}) > 0$.

The actual implementation of the Gibbs sampling technique in the specific example mentioned in the introduction is given in Section 4.4.

In Zeger and Karim (1991), the basic data consist of y_{ik}, the response for the kth unit in the ith stratum. In this way, no two-stage sampling is involved, thereby elliminating several steps in (i)-(v). However, Zeger and Karim (1991) allow the possibility of correlated errors ϵ_{ijk}, and thereafter put an inverse Wishart on the covariance matrix rather than the inverse gamma distribution as done in this chapter.

Zeger and Karim (1991) proposed modelling $h(\theta_{ijk})$, where h is a strictly monotone increasing function, by (4.2.2) rather than modelling θ_{ijk} itself. However, in their simulation work, they worked with the canonical link θ_{ijk}. Their calculations can be greatly simplified by adaptation of the Gilks-Wild algorithm.

Two special cases are of immense practical interest. First is the logistic regression model where $\theta_{ijk} = log(p_{ijk}/(1-p_{ijk}))$, p_{ijk} being the success probabilities in Bernoulli trials. Second is the log-linear model where $\theta_{ijk} = log(\xi_{ijk})$, ξ_{ijk} being Poisson means. We shall consider the second situation in Section 4.3.

The log-concavity idea is used slightly differently in Dellaportas and Smith (1993) where the prime objective is inference about $\boldsymbol{\beta}$ in generalized linear models and θ_{ijk}'s are modelled as functions of $\boldsymbol{\beta}$ without any error. Dellaportas and Smith (1993) have used a $\mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{D}_0)$ prior for $\boldsymbol{\beta}$ where $\boldsymbol{\beta}_0$ and $\boldsymbol{D}_0$ are known, and their method, unlike ours, does not use any unknown variance components.

Our method should also be contrasted to that of Albert (1988) which generalizes Leonard and Novick (1986). Albert's (1988) method when generalized to the present setting will first assign independent conjugate prior distributions

$$\pi(\theta_{ijk} \mid m_{ij}, \zeta) = exp[\zeta(m_{ij}\theta_{ijk} - \psi(\theta_{ijk})) + g(m_{ij}; \zeta)] \qquad (4.2.3)$$

to the θ_{ijk}. Next one assumes that $h(m_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ for some monotone function h. Subsequently, he assigns distributions (possibly diffuse) to the hyperparameters $\boldsymbol{\beta}$ and ζ. Thus, Albert's (1988) procedure amounts to modelling some function of

the prior mean through some linear model <u>without</u> any error. This can, of course, be generalized by adding an error component to the regression term. It should also be noted that Albert's paper was written before this recent surge of Monte Carlo integration. He, therefore, suggested approximations of the Bayes procedure by one or the other of three methods: (i) Laplace method (see e.g. Tierney and Kadane, 1986), (ii) quasi likelihood approach, and (iii) Brook's (1984) method. These approximations are, in general, unnecessary now with the advent of the sophisticated Monte Carlo integration techniques.

4.3 Analysis of Multi-Category Data

We now see how the results of the previous section help in the analysis of multi-category data. Consider m strata labelled $1, \cdots, m$. Within each stratum, several units are selected, and suppose that the responses of individuals within each selected unit are independent, and can be classified into J categories. For the kth selected unit within the ith stratum, let p_{ijk} denote the probability that an individual's response belongs to the jth category, and let Z_{ijk} denote the number of individuals whose response falls in the jth category ($j = 1, \cdots, J$; $k = 1, \cdots, n_i$). Then within the kth selected unit within the ith stratum, Z_{ijk} ($j = 1, \cdots, J$) has a joint multinomial $(n_i;\ p_{i1k}, \cdots, p_{iJk})$ distribution. Using the well-known relationship between multinomial and Poisson distributions $(Z_{i1k}, \cdots, Z_{iJk})$ has the same distribution as the joint conditional distribution of $(Y_{i1k}, \cdots, Y_{iJk})$ given $\sum_{j=1}^{J} Y_{ijk}$, where Y_{ijk} ($j = 1, \cdots, J$) are independent Poisson(ζ_{ijk}) and $p_{ijk} = \zeta_{ijk} / \sum_{j=1}^{J} \zeta_{ijk}$ ($j = 1, \cdots, J$). Thus, although the present structure is not strictly two-stage sampling within strata (since the suffix j corresponds to a category, and not a primary unit within the ith stratum), the results of the previous section apply (with $n_{ij} = n_i$ for all j and $c_i = J$ for all i) for finding the posterior distribution of ζ_{ijk}. The posterior means and variances

of p_{ijk} are simply obtained by using $p_{ijk} = \zeta_{ijk}/\sum_{j=1}^{J}\zeta_{ijk}$ $(j = 1, \cdots, J)$, and then using the Monte-Carlo integration algorithm.

To be specific, suppose that the Gibbs sampler uses t-iterates and G replications. The corresponding sampled ζ_{ijk} values are given by $\zeta_{ijkg}^{(t)}$ $(g = 1, \cdots, G)$. Then $E(p_{ijk} \mid \boldsymbol{y})$ is approximated by

$$G^{-1}\sum_{g=1}^{G}\frac{\zeta_{ijkg}^{(t)}}{\sum_{j=1}^{J}\zeta_{ijkg}^{(t)}},$$

while $E(p_{ijk}^2 \mid \boldsymbol{y})$ is approximated by

$$G^{-1}\sum_{g=1}^{G}\frac{(\zeta_{ijkg}^{(t)})^2}{(\sum_{j=1}^{J}\zeta_{ijkg}^{(t)})^2};$$

$V(p_{ijk} \mid \boldsymbol{y})$ is now approximated by using the individual approximations for $E(p_{ijk}^2 \mid \boldsymbol{y})$ and $E(p_{ijk} \mid \boldsymbol{y})$. Further, $E(p_{ijk}p_{i'j'k'} \mid \boldsymbol{y})$ is approximated by

$$G^{-1}\sum_{g=1}^{G}\left(\frac{\zeta_{ijkg}^{(t)}}{\sum_{j=1}^{J}\zeta_{ijkg}^{(t)}}\right)\left(\frac{\zeta_{i'j'k'g}^{(t)}}{\sum_{j=1}^{J}\zeta_{i'j'k'g}^{(t)}}\right)$$

which leads to an approximation for $Cov(p_{ijk}p_{i'j'k'} \mid \boldsymbol{y})$.

4.4 An Example

We illustrate the method of Sections 4.2 and 4.3 with an analysis of Canada Youth & AIDS Study data mentioned in the introduction. Recall that the question "how often have you had sexual intercourse?" had four response categories: never, once, a few times, and often. In this section we obtain the posterior mean and standard deviation of the proportion of Grade 9 students in the selected Province Newfoundland of Canada who would respond in any one of the four categories if sampled. No attempt is made to examine the question of reporting bias.

The different school boards are stratified by Catholic/Protestant and by urban/rural. This is an attempt to minimize the effect of selection bias, since some school boards refused to participate. Refusal was often based on the personal choice of the school board official and was related to how busy the school was, how many other issues the school had to deal with and a reticence to get involved in a situation perceived to have political ramifications. It is reasonable to assume that, within urban/rural and Catholic/Protestant categories within the geographical area studied here, student responses would be uncorrelated with reasons of refusal. We also assume that would-be responses are uncorrelated with student nonresponse (chiefly due to absence), though this is clearly a possible source of nonsampling error. Methods of modelling nonresponse in stratified sampling used by Stasny (1991) have not been developed for complex sampling designs.

School boards were selected according to a probability scheme where larger boards had a larger probability of being selected. Classes within schools within boards were randomly selected, and all students in attendance in the sampled classes were given the questionnaire. Let the stratum of Catholic/Protestant and urban/rural be indexed by r and c (rows and columns) respectively, where $r = 1, \cdots, R$ and $c = 1, \cdots, C$. Here in this example R=2 (Catholic and Protestant) and C=3 (Rural-Small Town, Town and Small City). For a given school board, it turns out that all schools within that school board fall within the same (r,c) stratum. Thus, the cluster is indexed by i corresponding to the school board, k corresponds to the school within a school board and j corresponding to the alternatives of the response. ($i = 1, \cdots, I,\ k = 1, \cdots, n_i,\ j = 1, \cdots, J$). Thus $n_{ij} = n_i$ for all j, and $c_i = J$ for all i.

We begin with the Poisson model for counts and then obtain the proportions as given in Section 4.3. The y-values within cluster (i, j, k) are distributed as

$$Y_{ijk} \overset{ind}{\sim} \text{Poisson}(\zeta_{ijk}) \tag{4.1.1}$$

Then, as discussed in Section 4.2, the natural parameter is modelled using the canonical link, specifically in the case of Poisson, the parameter $\theta_{ijk} = log\, \zeta_{ijk}$ will be modelled as

$$\theta_{ijk} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + u_{ij} + \epsilon_{ijk} \tag{4.4.5}$$

where

$$\begin{aligned} u_{ij} &\sim \mathrm{N}(0, \sigma_{\mathrm{B}}^2) \\ \epsilon_{ijk} &\sim \mathrm{N}(0, \sigma^2)\, . \end{aligned} \tag{4.4.6}$$

Keeping in mind that a given stratum i corresponds to a particular (r, c) combination, we have

$$\boldsymbol{x}_{ij}^T \boldsymbol{\beta} = \mu + \tau_j^J + \tau_r^R + \tau_c^C + \tau_{jr}^{JR} + \tau_{jc}^{JC} + \tau_{rc}^{RC} \tag{4.4.7}$$

$r = 1, \cdots, R$, $c = 1, \cdots, C$ and $j = 1, \cdots, J$. In the above μ is the general effect, τ_j^J is the main effect of the jth alternative of the response, τ_r^R is the main effect of the rth row, τ_c^C is the main effect of cth column, τ_{jr}^{JR} is the interaction effect of the jth response and the rth row, τ_{jc}^{JC} is the interaction effect of the jth response and the cth column and τ_{rc}^{RC} is the interaction effect of the rth row and the cth column. To avoid redundancy we assume the corner point restrictions namely

$$\tau_1^J = \tau_1^R = \tau_{1r}^{JR} = \tau_{j1}^{JR} = \tau_{1c}^{JC} = \tau_{j1}^{JC} = \tau_{r1}^{RC} = \tau_{1c}^{RC} = 0 \tag{4.4.8}$$

for all (r,c,j). The additive log-linear model (4.4.5) will cause estimates of the ζ_{ikj} to borrow strength from other estimates in board i and other estimates in school k. It is recommended in situations where some (i, j, k) cells have few samples, or even no samples.

Table 4.1 and 4.2 provides the hierarchical Bayes estimates and the associated standard errors for the proportion of students responding to the four categories in the forty selected schools within the fifteen school boards. Clearly, there is a distinction

between the sample proportions and the HB estimates. In particular, if no student responds for a specific category, for example "often", the sample proportions are clearly zeroes, whereas the HB method is usually assigning a small probability to the event. Judging the subjective nature of the response, the HB estimates are probably more meaningful than the sample proportions, at least for this category.

The biggest advantage of using the HB method instead of the sample proportions is the tremendous reduction in standard errors for all the three categories "Never", "Once" and "Few Times". For some of these categories, the reduction is often as high as fifty per cent. On the other hand, if no students respond to a certain category, based on the sample proportions, the estimated saturated standard errors are clearly zeroes. Such estimates are usually questionable, but the HB method is consistently rectifying this deficiency by producing positive estimates of these standard errors.

Perhaps the biggest advantage of the HB method lies in finite population sampling. If, after drawing a random sample, some clusters are not represented at all, the sample proportions for those clusters are not available. On the other hand, it is still possible to estimate the proportions in these categories by the HB method by borrowing strength from other clusters.

Table 4.1. Estimated and Sample Proportions

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
283	1	0.6684	0.1174	0.1868	0.0273
		(0.6072)	(0.1071)	(0.2857)	(0)
283	2	0.7195	0.1180	0.1349	0.0276
		(0.8696)	(0.1304)	(0)	(0)
283	3	0.6726	0.1203	0.1772	0.0299
		(0.5806)	(0.1290)	(0.2258)	(0.0646)
283	4	0.7053	0.1183	0.1472	0.0292
		(0.7600)	(0.1200)	(0.0800)	(0.0400)
283	5	0.7166	0.1118	0.1454	0.0262
		(0.7857)	(0.1071)	(0.1072)	(0)
283	6	0.7154	0.1049	0.1541	0.0256
		(0.7666)	(0.0667)	(0.1667)	(0)
284	1	0.5371	0.1218	0.2456	0.0954
		(0.2778)	(0.1111)	(0.3333)	(0.2778)
284	2	0.6771	0.0964	0.1619	0.0646
		(0.8148)	(0.0741)	(0.0370)	(0.0741)
284	3	0.6563	0.1029	0.1787	0.0622
		(0.8182)	(0.0909)	(0.0909)	(0)
284	4	0.6132	0.1011	0.2275	0.0582
		(0.5806)	(0.0968)	(0.3226)	(0)

Table 4.1. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
284	5	0.5678	0.1303	0.2263	0.0755
		(0.4074)	(0.2222)	(0.2593)	(0.1111)
284	6	0.6499	0.0992	0.1879	0.0630
		(0.8095)	(0.0476)	(0.1429)	(0)
285	1	0.3545	0.0531	0.4574	0.1350
		(0.2500)	(0.0833)	(0.5000)	(0.1667)
285	2	0.4339	0.0494	0.3982	0.1185
		(0.5416)	(0)	(0.4167)	(0.0417)
285	3	0.4331	0.0539	0.3731	0.1399
		(0.5000)	(0.0417)	(0.2916)	(0.1667)
287	1	0.5680	0.1128	0.2382	0.0810
		(0.6666)	(0.0556)	(0.1667)	(0.1111)
287	2	0.4798	0.1175	0.3174	0.0854
		(0.2963)	(0.1111)	(0.4444)	(0.1482)
287	3	0.5623	0.1246	0.2389	0.0742
		(0.5909)	(0.1818)	(0.1818)	(0.0455)
287	4	0.5214	0.1121	0.2936	0.0729
		(0.4583)	(0.0833)	(0.4167)	(0.0417)
287	5	0.6332	0.1067	0.1937	0.0664
		(0.9091)	(0.0909)	(0)	(0)

Table 4.1. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
291	1	0.6431	0.1096	0.1987	0.0486
		(0.6956)	(0.0870)	(0.2174)	(0)
291	2	0.6495	0.1129	0.1871	0.0505
		(0.6333)	(0.1333)	(0.1667)	(0.0667)
292	1	0.5297	0.1711	0.1368	0.1624
		(0.6250)	(0)	(0)	(0.3750)
292	2	0.5405	0.1901	0.1392	0.1302
		(0.5000)	(0.2778)	(0.1666)	(0.0556)
293	1	0.6916	0.0715	0.1766	0.0602
		(0.6667)	(0.0833)	(0.1667)	(0.0833)
293	2	0.7321	0.0588	0.1594	0.0496
		(0.7812)	(0.0313)	(0.1562)	(0.0313)
293	3	0.6775	0.0755	0.1865	0.0606
		(0.6500)	(0.1000)	(0.2000)	(0.0500)
295	1	0.4916	0.1320	0.2167	0.1596
		(0.4865)	(0.1351)	(0.2162)	(0.1622)
297	1	0.4049	0.0879	0.3547	0.1525
		(0.4000)	(0.0800)	(0.3600)	(0.1600)
301	1	0.4358	0.1516	0.3012	0.1114
		(0.4211)	(0.1578)	(0.3158)	(0.1053)

Table 4.1. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
305	1	0.6156	0.1489	0.1692	0.0663
		(0.6785)	(0.1786)	(0.1429)	(0)
305	2	0.5985	0.1449	0.1722	0.0845
		(0.6667)	(0.0952)	(0.0952)	(0.1429)
305	3	0.5062	0.1681	0.2371	0.0887
		(0.2941)	(0.1765)	(0.4118)	(0.1176)
306	1	0.7005	0.0601	0.1685	0.0709
		(0.6250)	(0.0625)	(0.2500)	(0.0625)
306	2	0.7703	0.0472	0.1244	0.0581
		(0.8000)	(0.0333)	(0.1000)	(0.0667)
308	1	0.8160	0.0525	0.0593	0.0721
		(0.8158)	(0.0526)	(0.0527)	(0.0789)
311	1	0.5572	0.1382	0.1361	0.1685
		(0.5334)	(0.1333)	(0.1333)	(0.2000)
314	1	0.8032	0.0724	0.1055	0.0189
		(0.8064)	(0.0968)	(0.0968)	(0)
314	2	0.7838	0.0748	0.1197	0.0217
		(0.7408)	(0.0741)	(0.1481)	(0.0370)
314	3	0.7898	0.0745	0.1140	0.0217
		(0.8572)	(0.0476)	(0.0952)	(0)

Table 4.2. Standard Errors for Estimated and Sample Proportions

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
283	1	0.0570	0.0348	0.0479	0.0136
		(0.0923)	(0.0584)	(0.0854)	(0)
283	2	0.0531	0.0349	0.0382	0.0138
		(0.0702)	(0.0702)	(0)	(0)
283	3	0.0547	0.0352	0.0435	0.0148
		(0.0886)	(0.0602)	(0.0751)	(0.0442)
283	4	0.0527	0.0349	0.0382	0.0144
		(0.0854)	(0.0650)	(0.0543)	(0.0392)
283	5	0.0522	0.0332	0.0375	0.0130
		(0.0775)	(0.0584)	(0.0585)	(0)
283	6	0.0518	0.0320	0.0379	0.0129
		(0.0772)	(0.0456)	(0.0680)	(0)
284	1	0.0818	0.0402	0.0598	0.0402
		(0.1056)	(0.0741)	(0.1111)	(0.1056)
284	2	0.0610	0.0305	0.0490	0.0234
		(0.0748)	(0.0504)	(0.0363)	(0.0504)
284	3	0.0588	0.0327	0.0474	0.0227
		(0.0822)	(0.0613)	(0.0613)	(0)
284	4	0.0575	0.0314	0.0518	0.0213
		(0.0886)	(0.0531)	(0.0840)	(0)

Table 4.2. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
284	5	0.0698	0.0436	0.0524	0.0283
		(0.0946)	(0.0800)	(0.0843)	(0.0605)
284	6	0.0590	0.0317	0.0474	0.0234
		(0.0857)	(0.0465)	(0.0764)	(0)
285	1	0.0686	0.0252	0.0711	0.0417
		(0.0722)	(0.0461)	(0.0833)	(0.0621)
285	2	0.0695	0.0236	0.0666	0.0393
		(0.1017)	(0)	(0.1006)	(0.0408)
285	3	0.0705	0.0261	0.0710	0.0445
		(0.1021)	(0.0408)	(0.0928)	(0.0761)
287	1	0.0633	0.0376	0.0541	0.0313
		(0.1111)	(0.0540)	(0.0878)	(0.0741)
287	2	0.0792	0.0380	0.0720	0.0335
		(0.0879)	(0.0605)	(0.0956)	(0.0684)
287	3	0.0638	0.0404	0.0536	0.0285
		(0.1048)	(0.0822)	(0.0822)	(0.0444)
287	4	0.0669	0.0367	0.0636	0.0279
		(0.1017)	(0.0564)	(0.1006)	(0.0408)
287	5	0.0725	0.0354	0.0572	0.0257
		(0.0613)	(0.0613)	(0)	(0)

Table 4.2. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
291	1	0.0717	0.0427	0.0598	0.0296
		(0.0959)	(0.0588)	(0.0860)	(0)
291	2	0.0693	0.0429	0.0583	0.0301
		(0.0880)	(0.0621)	(0.0680)	(0.0456)
292	1	0.1001	0.0696	0.0701	0.0742
		(0.1712)	(0)	(0)	(0.1712)
292	2	0.0958	0.0746	0.0695	0.0595
		(0.1179)	(0.1056)	(0.0878)	(0.0540)
293	1	0.0619	0.0307	0.0503	0.0288
		(0.0962)	(0.0564)	(0.0761)	(0.0564)
293	2	0.0573	0.0257	0.0460	0.0237
		(0.0731)	(0.0308)	(0.0641)	(0.0308)
293	3	0.0656	0.0333	0.0536	0.0291
		(0.1067)	(0.0671)	(0.0894)	(0.0487)
295	1	0.0768	0.0526	0.0635	0.0559
		(0.0822)	(0.0562)	(0.0677)	(0.0606)
297	1	0.0973	0.0502	0.0896	0.0680
		(0.0980)	(0.0543)	(0.0960)	(0.0733)
301	1	0.1006	0.0692	0.0960	0.0639
		(0.1133)	(0.0836)	(0.1066)	(0.0704)

Table 4.2. (continued)

BOARD	CLASS	NEVER	ONCE	FEW TIMES	OFTEN
305	1	0.0703	0.0485	0.0510	0.0308
		(0.0883)	(0.0724)	(0.0661)	(0)
305	2	0.0722	0.0496	0.0527	0.0378
		(0.1029)	(0.0640)	(0.0640)	(0.0764)
305	3	0.0874	0.0546	0.0764	0.0395
		(0.1105)	(0.0925)	(0.1194)	(0.0781)
306	1	0.0806	0.0366	0.0675	0.0390
		(0.1210)	(0.0605)	(0.1083)	(0.0605)
306	2	0.0597	0.0291	0.0477	0.0315
		(0.0730)	(0.0328)	(0.0548)	(0.0456)
308	1	0.0583	0.0335	0.0354	0.0378
		(0.0629)	(0.0362)	(0.0362)	(0.0437)
311	1	0.1167	0.0791	0.0792	0.0895
		(0.1288)	(0.0878)	(0.0878)	(0.1033)
314	1	0.0484	0.0309	0.0368	0.0134
		(0.0710)	(0.0531)	(0.0531)	(0)
314	2	0.0533	0.0319	0.0415	0.0154
		(0.0843)	(0.0504)	(0.0684)	(0.0363)
314	3	0.0521	0.0322	0.0399	0.0156
		(0.0763)	(0.0465)	(0.0640)	(0)

CHAPTER 5
SUMMARY AND FUTURE RESEARCH

5.1 Summary

In this dissertation, we have considered several problems where the hierarchical Bayes (HB) methodology is used to obtain estimates and the associated standard errors. The Bayesian methodology has been applied to two specific problems of small area estimation, namely, the adjustment of the census undercount and categorical survey data. We have also provided a hierarchical Bayes refinement of Hoadley's Quality Measurement Plan (QMP).

The hierarchical Bayes procedure proposed in Chapter 2 for the adjustment of the 1990 Census undercount overcomes many of the criticisms levelled against the Bayesian procedures of earlier authors. In particular, we have dicussed a model-based approach which eliminates the need for assuming variance-covariance matrices of the adjustment factors to be known, which was hitherto assumed known in any Bayesian or non-Bayesian analysis.

Despite its wide publicity, the QMP developed by Hoadley (1981) has been criticized by many statisticians. One of the main criticisms levelled against the procedure is that the procedure is heuristic and it would require high-dimensional numerical integration for a full Bayesian implementation. In Chapter 3, we have provided a HB procedure that will avoid all the ad hoc approximations needed in Hoadley's solution.

In Chapter 4, a full Bayesian analysis is provided for categorical survey data, where data are classified into multi-(not necessarily two) categories. More generally, we have provided a complete HB analysis for two-stage sampling within strata based on generalized linear models. The technique has been used to produce estimates and standard errors in the Canada youth and AIDS study data.

In all chapters of this dissertation, the implementation of the HB methodology have been illustrated by adopting a Monte Carlo integration technique known as the Gibbs sampler. Using this procedure, the posterior density as well as conditional mean and variance can be obtained with considerable ease. Also, a special technique called the adaptive rejection sampling has been extensively used to generate samples from log-concave densities.

5.2 Future Research

The Gibbs sampler and iterative simulation methods are potentially very helpful for summarizing univariate and multivariate distributions. In all of our applications, we have employed a single sequence of $t \times G$ Gibbs iterates, storing every t^{th} iterate to provide i.i.d *k-tuples* $(U_{1g}^{(t)}, \cdots, U_{kg}^{(t)})$, $(g = 1, \cdots, G)$. Since there are no proper established techniques to monitor convergence of an iterative simulation, we have employed crude existing techniques for assessing convergence. But it is possible that when using a single sequence, the inferences may be unduly influenced by slow-moving realizations of the iterative simulation. It is important to establish the convergence by implementing quantitative methods in monitoring convergence. One such possible strategy is to use several independent sequences, with starting points sampled from an overdispersed distribution, as recommended by Gelman and Rubin (1992). Also, in the case of simulating samples from non-logconcave densities, it is possible to use the Adaptive Rejection Metropolis sampling as in Gilks et al. (1993).

Coming to the specific problems considered in this dissertation, in the adjustment of census undercount, we have not taken into account the pairwise correlation of the adjustment factors between the different poststrata, since the sample correlations were too small compared to the variances. We have, in a previous study, considered a general correlation structure by assuming Wishart type priors on the variance-covariance matrix, but this yielded unreasonable estimates. The case in which special type of correlation structure is more appropriate needs further investigation.

In addition to the refinement of Hoadley's QMP, it is important to investigate the possibility of a full Bayesian implementation in the additive and multiplicative model proposed by Irony et al.(1992) for analyzing discrete time-series for quality data.

We can extend the HB analysis of categorical survey data to prediction in the case of finite population sampling. As discussed in Chapter 4, the HB method is well suited for predictive inference, since the method estimates the unsampled portion by borrowing strength from related areas.

BIBLIOGRAPHY

Albert, J. H. (1988). Computational methods using a Bayesian hierarchical generalized linear model. *J. Amer. Statist. Assoc.*, **83**, 1037-1044.

Angers, J. F. and Berger, J. O. (1991). Robust hierarchical Bayes estimation of exchangeable means. *Canadian J. Statist.***19**, 39-56.

Barlow, R. E. and Irony, T. Z. (1992) Foundations of statistical quality control. *Current Issues in Statistical Inference: Essays in Honor of D. Basu.* IMS Lecture Notes- Monograph Series. Eds. M. Ghosh and P. K. Pathak 99-112.

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error–components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28–36.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition. Springer-Verlag, New York.

Brooks, R. J. (1984). Approximate likelihood ratio tests in the analysis of beta-binomial data. *Applied Statistics*, **33**, 285-289.

Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics*, **19**, 1748–1770.

Datta, G. S., Ghosh, M., Huang, E. T., Isaki, C. T., Schultz, L. K. and Tsay, J. H. (1992). Hierarchical and empirical Bayes methods for adjustment of census undercount: The 1988 Missouri Dress Rehearsal Data. *Survey Methodology*, **18**, 95-108.

Datta, G. S. and Lahiri, P. (1992). Robust hierarchical Bayes estimation of small area characteristics in presence of covariates. Technical Report No. 92-28, Dept. of Statistics, University of Georgia, Athens.

Dellaportas, P. and Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics*, **42**, 443-459.

Ericksen, E.P. and Kadane, J.B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.*, **80**, 98–131.

Ericksen, E.P. and Kadane, J.B. (1987). Sensitivity analysis of local estimates of undercount in the 1980 U.S. Census. *Small Area Statistics*. Eds. R. Platek, J.N.K., Rao, C.E., Sarndal, and M.P. Singh. Wiley, New York, pp. 23–45.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269–277.

Fienberg, S. E. (1992). An adjusted census in 1990? The trial. *Chance, New Directions for Stat. and Computers*, **5**, 28-38.

Freedman, D. A. and Navidi, W. C. (1986). Regression models for adjusting the 1980 census (with discussion). *Statistical Science*, **1**, 3-39.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics*, **29**, 1-15.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.*, **85**, 972-985.

Gelfand. A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398-409.

Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Commun. in Statist.-Theory Meth.*, **20**(5 & 6), 1747-1766.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

George, E. I., Makov, U. E. and Smith, A. F. M. (1993). Conjugate Likelihood Distribution, To appear in the *Scandinavian Journal of Statistics*.

Ghosh, M. (1992). Hierarchical and empirical Bayes multivariate estimation. *Current Issues in Statistical Inference: Essays in Honor of D. Basu.* IMS Lecture Notes- Monograph Series. Eds. M. Ghosh and P. K. Pathak. **17**, pp. 151-177.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.*, **82**, 1153–1162.

Ghosh, M. and Lahiri, P. (1992). A hierarchical Bayes approach to small area estimation with auxillary information. *Bayesian Analysis in Statistics and Econometrics*, Eds. P. K. Goel and N. S. Iyengar, Lecture Notes in Statistics, **75**. Springer-Verlag, New York. pp. 107-125.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.*, **74**, 269–277.

Ghosh, M. and Rao, J.N.K. (1991). Small area estimation: An appraisal. Technical Report 390, Dept. Statistics, Univ. Florida, Gainesville.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1993). Adaptive rejection metropolis sampling for Gibbs sampling. *Submitted for review.*

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337-348.

Good, I. J. (1965). *The Estimation of Probabilities, An Essay on Modern Bayesian Methods.* M.I.T. Press, Cambridge, Massachusetts.

Hoadley, B. (1981). The Quality Measurement Plan (QMP). *The Bell System Technical Journal*, **60**, 215-273.

Irony, T. Z., Pereira, C. A. de B., and Barlow, R. E. (1992). *Bayesian Statistics*, 4, Eds. J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith. Oxford University Press, New York, pp. 675-688.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models), *J. Amer. Statist. Assoc.*, **84**, 717-726.

Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.*, **84**, 710-716.

King, A. J. C., Beazley, R. P., Warren, W. K., Hankins, C. A., Robertson, A. S. and Radford, J. L. (1988). *Canada Youth & AIDS Study.* Federal Centre for AIDS, Ottawa.

Leonard, T. and Novick, M. R. (1986). Bayesian full rank marginalization for two-way contingency tables. *J. of Educ. Stat.*, **11**, 33-56.

Lindley, D. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *J. R. Statist. Soc. B*, **34**, 1-41.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (2nd ed.).* Chapman & Hall, London.

Morris, C. (1981). Parametric empirical Bayes confidence intervals, *Scientific Inference, Data Analysis, and Robustness*, Eds. G. E. P. Box, T. Leonard and C. F. Jeff Wu, Academic Press, pp. 25-50.

Morris, C. (1983). Parametric empirical Bayes inference and applications, *J. Amer. Statist. Assoc.*, **78**, 47-65.

Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared errors of small–area estimators. *J. Amer. Statist. Assoc.*, **85**, 163–171.

Stasny, E. A. (1991). Hierarchical models for victimization and nonresponse in the National Crime Survey. *J. Amer. Statist. Assoc.*, **86**, 296-303.

Stroud, T. W. F. (1991). Hierarchical Bayes predictive means and variances with application to sample survey inference. *Commun. in Statist.-Theory Meth.*, **20**(1), 13-36.

Stroud, T. W. F. (1992). Bayesian inference from categorical survey data. Preprint No. 1991-3, Dept. of Mathematics and Statistics, Queen's University.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82-86.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; A Gibbs sampling approach. *J. Amer. Statist. Assoc.*, **86**, 79-86.

BIOGRAPHICAL SKETCH

Kannan Natarajan was born on October 21, 1966, in Madras, Tamil Nadu, India. He received his Bachelor of Science degree with major in statistics in 1987 and thereafter joined the Indian Statistical Institute, Calcutta, India, where he received his Master of Statistics degree in 1989. In August 1989, he joined the doctoral program in statistics in the University of Florida at Gainesville. He expects to receive a Ph.D. degree in December 1993. During his time at the University of Florida, he worked as a research assistant to Dr. Malay Ghosh. He was also employed as a teaching assistant in the Department of Statistics. Upon graduation, he will join the Division of Clinical Statistics and Computing, Abbott Laboratories, Abbott Park.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh

Malay Ghosh, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Alan Agresti
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard L. Scheaffer
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ramon C. Littell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Patrick A. Thompson
Assistant Professor of Decision
and Information Sciences

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1993

Dean, Graduate School